

STIC-Biotech/ChemLib

143 104

my

From: Duffy, Patricia  
Sent: Tuesday, January 25, 2005 7:26 AM  
To: STIC-Biotech/ChemLib  
Subject: protein sequence search 09/993,292

In re: 09/993,292

Please search SEQ ID NO:24.

Please search in commercial and interference databases.

Please print out top 50 hits.

Thanks.

*Patricia A. Duffy, Ph.D.*

*Art Unit 1645, Remsen 3B05*

*571-272-0855*

RECEIVED  
JAN 25 2005  
STIC

\*\*\*\*\*

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2- \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA Sequence: # \_\_\_\_\_  
AA Sequence :# \_\_\_\_\_  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**This Page Blank (uspto)**



From: Chan, Christina  
Sent: Tuesday, January 04, 2005 4:14 PM  
To: Duffy, Patricia; STIC-Biotech/ChemLib  
Subject: RE: rush sequence search

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644  
(571)-272-0841  
Remsen, 3E89

-----Original Message-----

From: Duffy, Patricia  
Sent: Tuesday, January 04, 2005 4:08 PM  
To: Chan, Christina  
Subject: rush sequence search  
Importance: High

Christina,

Please approve rush sequence request. Amendment due this biweek.

IN re: 09/993,292

Please search protein sequences of SEQ ID NOS:2 and 28.  
Please include both an interference and commercial database search.  
Please print out top 50 hits in each category.

Thank you.

Patricia A. Duffy, Ph.D.  
Art Unit 1645, Remsen 3B05  
571-272-0855

AA  
2-305  
28-303

\*\*\*\*\*  
STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2- \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*  
Type of Search

NA Sequence: # \_\_\_\_\_  
AA Sequence: # \_\_\_\_\_  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*  
Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

Date completed: \_\_\_\_\_

Searcher: Beverly e 2528

Terminal time: \_\_\_\_\_

Elapsed time: \_\_\_\_\_

CPU time: \_\_\_\_\_

Total time: \_\_\_\_\_

Number of Searches: \_\_\_\_\_

Number of Databases: \_\_\_\_\_

Search Site

\_\_\_\_\_ STIC

\_\_\_\_\_ CM-1

\_\_\_\_\_ Pre-S

Type of Search

\_\_\_\_\_ N.A. Sequence

\_\_\_\_\_ A.A. Sequence

\_\_\_\_\_ Structure

\_\_\_\_\_ Bibliographic

Vendors

\_\_\_\_\_ IG

\_\_\_\_\_ STN

\_\_\_\_\_ Dialog

\_\_\_\_\_ APS

\_\_\_\_\_ Geninfo

\_\_\_\_\_ SDC

\_\_\_\_\_ DARC/Questel

\_\_\_\_\_ Other CGN

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:07:40 ; Search time 39 Seconds  
(without alignments)  
515.240 Million cell updates/sec

Title: US-09-993-292B-24

Perfect score: 1510

Sequence: 1 MTGIFAEQTEVVKSAIETA.....TCNEYQQRHGKKTLLLEVPDI 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 1376  | 91.1        | 309    | 1  | US-08-557-115-3      |
| 2          | 1376  | 91.1        | 309    | 5  | PCT-US94-05869-3     |
| 3          | 111.5 | 7.4         | 1036   | 4  | US-09-543-681A-7736  |
| 4          | 109   | 7.2         | 515    | 4  | US-09-107-532A-5317  |
| 5          | 108.5 | 7.2         | 1196   | 4  | US-09-107-532A-3944  |
| 6          | 106.5 | 7.1         | 3696   | 3  | US-09-134-001C-5080  |
| 7          | 106   | 7.0         | 1231   | 4  | US-09-107-532A-5150  |
| 8          | 104   | 6.9         | 815    | 4  | US-09-328-352-4284   |
| 9          | 104   | 6.9         | 1211   | 3  | US-09-134-001C-4820  |
| 10         | 103   | 6.8         | 284    | 4  | US-09-134-000C-3594  |
| 11         | 103   | 6.8         | 885    | 4  | US-09-710-279-1660   |
| 12         | 103   | 6.8         | 961    | 4  | US-09-914-259-66     |
| 13         | 102   | 6.8         | 722    | 4  | US-09-248-796A-20613 |
| 14         | 101.5 | 6.7         | 1857   | 4  | US-09-917-254-91     |
| 15         | 101.5 | 6.7         | 1972   | 4  | US-09-538-092-1084   |
| 16         | 101   | 6.7         | 1129   | 4  | US-09-543-681A-8019  |
| 17         | 100.5 | 6.7         | 526    | 4  | US-09-447-497-15     |
| 18         | 100.5 | 6.7         | 553    | 4  | US-09-447-497-14     |
| 19         | 100.5 | 6.7         | 566    | 4  | US-09-538-092-807    |
| 20         | 100.5 | 6.7         | 609    | 4  | US-09-447-497-12     |
| 21         | 100.5 | 6.7         | 613    | 4  | US-09-447-497-9      |
| 22         | 100.5 | 6.7         | 623    | 4  | US-09-447-497-11     |
| 23         | 100.5 | 6.7         | 627    | 4  | US-09-447-497-8      |
| 24         | 100   | 6.6         | 1886   | 3  | US-08-938-105-3      |
| 25         | 100   | 6.6         | 1939   | 3  | US-09-310-187A-1     |
| 26         | 100   | 6.6         | 1939   | 4  | US-09-538-092-917    |
| 27         | 99.5  | 6.6         | 924    | 4  | US-09-248-796A-18798 |

|     |      |     |       |   |                      |                   |
|-----|------|-----|-------|---|----------------------|-------------------|
| 28  | 99.5 | 6.6 | 1066  | 3 | US-09-541-782-8      | Sequence 8, Appli |
| 29  | 99.5 | 6.6 | 1066  | 4 | US-09-723-820-8      | Sequence 8, Appli |
| 30  | 99.5 | 6.6 | 1066  | 4 | US-10-270-085-8      | Sequence 8, Appli |
| 31  | 99.5 | 6.6 | 2349  | 4 | US-09-538-092-914    | Sequence 914, App |
| 32  | 98   | 6.5 | 618   | 4 | US-09-134-000C-4522  | Sequence 4522, Ap |
| 33  | 98   | 6.5 | 1122  | 4 | US-09-489-039A-8554  | Sequence 8554, Ap |
| 34  | 97.5 | 6.5 | 718   | 4 | US-09-540-236-2753   | Sequence 2753, Ap |
| 35  | 97.5 | 6.5 | 804   | 1 | US-08-785-428-2      | Sequence 2, Appli |
| 36  | 97.5 | 6.5 | 804   | 2 | US-08-996-797-2      | Sequence 2, Appli |
| 37  | 97.5 | 6.5 | 1010  | 3 | US-09-134-001C-5178  | Sequence 5178, Ap |
| 38  | 97   | 6.4 | 975   | 4 | US-09-914-259-19     | Sequence 19, Appl |
| 39  | 96.5 | 6.4 | 317   | 4 | US-09-248-796A-19212 | Sequence 19212, A |
| 40  | 96.5 | 6.4 | 858   | 4 | US-09-248-796A-19055 | Sequence 19055, A |
| 41  | 96.5 | 6.4 | 1312  | 2 | US-08-592-126-148    | Sequence 148, App |
| 42  | 96.5 | 6.4 | 1312  | 2 | US-08-687-080-51     | Sequence 51, Appl |
| 43  | 96.5 | 6.4 | 1312  | 4 | US-09-168-595-148    | Sequence 148, App |
| 44  | 96   | 6.4 | 360   | 4 | US-09-248-796A-18018 | Sequence 18018, A |
| 45  | 96   | 6.4 | 1086  | 4 | US-09-543-681A-7696  | Sequence 7696, Ap |
| 46  | 96   | 6.4 | 1960  | 4 | US-09-538-092-1077   | Sequence 1077, Ap |
| 47  | 96   | 6.4 | 2871  | 4 | US-09-538-092-936    | Sequence 936, App |
| 48  | 95.5 | 6.3 | 1972  | 4 | US-08-875-435B-3     | Sequence 3, Appli |
| 49  | 94.5 | 6.3 | 746   | 3 | US-08-434-000A-4     | Sequence 4, Appli |
| 50  | 94.5 | 6.3 | 746   | 3 | US-09-312-157-4      | Sequence 4, Appli |
| 51  | 94.5 | 6.3 | 815   | 4 | US-09-914-259-18     | Sequence 18, Appl |
| 52  | 94.5 | 6.3 | 1454  | 4 | US-09-328-352-5793   | Sequence 5793, Ap |
| 53  | 94.5 | 6.3 | 1976  | 4 | US-09-538-092-1078   | Sequence 1078, Ap |
| 54  | 94   | 6.2 | 606   | 4 | US-08-477-831C-2     | Sequence 2, Appli |
| 55  | 94   | 6.2 | 631   | 4 | US-08-477-831C-11    | Sequence 11, Appl |
| 56  | 94   | 6.2 | 3878  | 4 | US-09-914-259-11     | Sequence 11, Appl |
| 57  | 93.5 | 6.2 | 643   | 4 | US-09-538-092-844    | Sequence 844, App |
| 58  | 93.5 | 6.2 | 733   | 4 | US-09-328-352-5599   | Sequence 5599, Ap |
| 59  | 93.5 | 6.2 | 829   | 4 | US-09-248-796A-20145 | Sequence 20145, A |
| 60  | 93.5 | 6.2 | 1032  | 4 | US-09-914-259-26     | Sequence 26, Appl |
| 61  | 93.5 | 6.2 | 1032  | 4 | US-09-538-092-1293   | Sequence 1293, Ap |
| 62  | 93   | 6.2 | 451   | 4 | US-09-134-000C-3849  | Sequence 3849, Ap |
| 63  | 93   | 6.2 | 885   | 2 | US-08-533-306A-4     | Sequence 4, Appli |
| 64  | 93   | 6.2 | 885   | 2 | US-08-742-923A-4     | Sequence 4, Appli |
| 65  | 93   | 6.2 | 10182 | 3 | US-09-134-001C-3159  | Sequence 3159, Ap |
| 66  | 92.5 | 6.1 | 1027  | 4 | US-09-914-259-27     | Sequence 27, Appl |
| 67  | 92.5 | 6.1 | 1111  | 1 | US-08-317-450B-15    | Sequence 15, Appl |
| 68  | 92.5 | 6.1 | 1111  | 3 | US-08-800-593-15     | Sequence 15, Appl |
| 69  | 92.5 | 6.1 | 1172  | 4 | US-09-560-385A-28    | Sequence 28, Appl |
| 70  | 92.5 | 6.1 | 1172  | 4 | US-09-560-385A-32    | Sequence 32, Appl |
| 71  | 92.5 | 6.1 | 1193  | 1 | US-08-317-450B-13    | Sequence 13, Appl |
| 72  | 92.5 | 6.1 | 1193  | 3 | US-08-800-593-13     | Sequence 13, Appl |
| 73  | 92.5 | 6.1 | 1193  | 4 | US-09-560-385A-26    | Sequence 26, Appl |
| 74  | 92.5 | 6.1 | 1193  | 4 | US-09-560-385A-30    | Sequence 30, Appl |
| 75  | 92.5 | 6.1 | 2482  | 1 | US-08-328-254-6      | Sequence 6, Appli |
| 76  | 92.5 | 6.1 | 3075  | 2 | US-08-460-309-5      | Sequence 5, Appli |
| 77  | 92.5 | 6.1 | 3075  | 2 | US-08-125-077-5      | Sequence 5, Appli |
| 78  | 92.5 | 6.1 | 3210  | 4 | US-09-538-092-1154   | Sequence 1154, Ap |
| 79  | 92.5 | 6.1 | 3248  | 1 | US-08-353-700-1      | Sequence 1, Appli |
| 80  | 92.5 | 6.1 | 3248  | 5 | PCT-US95-16216-1     | Sequence 1, Appli |
| 81  | 92   | 6.1 | 440   | 4 | US-08-302-756E-35    | Sequence 35, Appl |
| 82  | 92   | 6.1 | 477   | 1 | US-08-402-217A-3     | Sequence 3, Appli |
| 83  | 92   | 6.1 | 477   | 1 | US-08-700-178-3      | Sequence 3, Appli |
| 84  | 92   | 6.1 | 477   | 3 | US-08-995-654-3      | Sequence 3, Appli |
| 85  | 92   | 6.1 | 849   | 3 | US-09-157-257-4      | Sequence 4, Appli |
| 86  | 92   | 6.1 | 5024  | 4 | US-09-710-279-2964   | Sequence 2964, Ap |
| 87  | 91.5 | 6.1 | 350   | 4 | US-09-489-039A-8352  | Sequence 8352, Ap |
| 88  | 91.5 | 6.1 | 652   | 4 | US-09-489-039A-8360  | Sequence 3360, Ap |
| 89  | 91.5 | 6.1 | 1939  | 4 | US-09-538-092-915    | Sequence 915, App |
| 90  | 91.5 | 6.1 | 2285  | 3 | US-09-308-375-2      | Sequence 2, Appli |
| 91  | 91   | 6.0 | 756   | 4 | US-09-248-796A-17753 | Sequence 17753, A |
| 92  | 91   | 6.0 | 2383  | 4 | US-09-492-709A-302   | Sequence 302, App |
| 93  | 90.5 | 6.0 | 451   | 4 | US-09-538-092-706    | Sequence 706, App |
| 94  | 90.5 | 6.0 | 574   | 3 | US-09-134-001C-3382  | Sequence 3382, Ap |
| 95  | 90.5 | 6.0 | 1937  | 4 | US-09-538-092-918    | Sequence 918, App |
| 96  | 90   | 6.0 | 346   | 4 | US-08-887-534A-68    | Sequence 68, Appl |
| 97  | 90   | 6.0 | 346   | 4 | US-09-527-431-68     | Sequence 68, Appl |
| 98  | 90   | 6.0 | 346   | 4 | US-09-446-861-68     | Sequence 68, Appl |
| 99  | 90   | 6.0 | 883   | 4 | US-09-489-039A-12755 | Sequence 12755, A |
| 100 | 90   | 6.0 | 1427  | 4 | US-09-538-092-1044   | Sequence 1044, Ap |

## ALIGNMENTS

RESULT 1  
US-08-557-115-3  
; Sequence 3, Application US/08557115  
; Patent No. 5731151  
; GENERAL INFORMATION:  
; APPLICANT: King, Harold C.  
; APPLICANT: Sathish, Mundayoor  
; APPLICANT: Shinnick, Thomas M.  
; TITLE OF INVENTION: Regulator of Contact-Mediated Hemolysin  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jones & Askew  
; STREET: 191 Peachtree Street, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.A.  
; ZIP: 30303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/557,115  
; FILING DATE: 26-JUN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greene, Jamie L.  
; REGISTRATION NUMBER: 32,467  
; REFERENCE/DOCKET NUMBER: 03063-0171US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 818-3700  
; TELEFAX: (404) 818-3799  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 309 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US-08-557-115-3

Query Match 91.1%; Score 1376; DB 1; Length 309;  
Best Local Similarity 90.4%; Pred. No. 2.6e-132;  
Matches 272; Conservative 16; Mismatches 13; Indels 0; Gaps 0;  
QY 1 MTGIFAQTVVVKSAIETADGALDFYNNKYLDQVLPWKTFDETIKELSRFKQEYSQASV 60  
DB 1 MTEIVADKTVEVVKNAIETADGALDLYNNKYLDQVLPWKTFDETIKELSRFKQEYSQASV 60  
QY 61 LVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRIL 120  
DB 61 LVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRIL 120  
QY 121 DGVNKLNEAOKSLGSSQFNNSAGKLLALDSQLTNDFSEKSSYFQSQVDRIKEAYAG 180  
DB 121 DGIITKLENAOKSLVSSQFNNSAGKLLALDSQLTNDFSEKSSYFQSQVDRIKEAYAG 180  
QY 181 AAAGVAGPFGILIIYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDAAK 240  
DB 181 AAAGVAGPFGILIIYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDAAK 240  
QY 241 LKLAETIAAIGETETETTRFYVDYDDLMLSLKGAACKMINTCNEYQQRHGKKTLLFV 300  
DB 241 LKLTETIAAIGETETETTRFYVDYDDLMLSLKGAACKMINTCNEYQQRHGKKTLLFV 300

QY 301 P 301  
DB 301 P 301  
RESULT 2  
PCT-US94-05869-3  
; Sequence 3, Application PC/TUS9405869  
; GENERAL INFORMATION:  
; APPLICANT: King, C. H.  
; APPLICANT: Sathish, Mundayoor  
; APPLICANT: Shinnick, Thomas M.  
; TITLE OF INVENTION: REGULATOR OF CONTACT-MEDIATED HEMOLYSIN  
; TITLE OF INVENTION: OF MYCOBACTERIUM TUBERCULOSIS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Suite 1200, The Candler Building  
; STREET: 127 Peachtree Street, N.E.  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303-1811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/05869  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perryman, David G.  
; REGISTRATION NUMBER: 33,438  
; REFERENCE/DOCKET NUMBER: 1414.611  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 688-0770  
; TELEFAX: (404) 688-9880  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 309 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US94-05869-3

Query Match 91.1%; Score 1376; DB 5; Length 309;  
Best Local Similarity 90.4%; Pred. No. 2.6e-132;  
Matches 272; Conservative 16; Mismatches 13; Indels 0; Gaps 0;  
QY 1 MTGIFAQTVVVKSAIETADGALDFYNNKYLDQVLPWKTFDETIKELSRFKQEYSQASV 60  
DB 1 MTEIVADKTVEVVKNAIETADGALDLYNNKYLDQVLPWKTFDETIKELSRFKQEYSQASV 60  
QY 61 LVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRIL 120  
DB 61 LVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRIL 120  
QY 121 DGVNKLNEAOKSLGSSQFNNSAGKLLALDSQLTNDFSEKSSYFQSQVDRIKEAYAG 180  
DB 121 DGIITKLENAOKSLVSSQFNNSAGKLLALDSQLTNDFSEKSSYFQSQVDRIKEAYAG 180  
QY 181 AAAGVAGPFGILIIYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDAAK 240  
DB 181 AAAGVAGPFGILIIYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDAAK 240  
QY 241 LKLAETIAAIGETETETTRFYVDYDDLMLSLKGAACKMINTCNEYQQRHGKKTLLFV 300  
DB 241 LKLTETIAAIGETETETTRFYVDYDDLMLSLKGAACKMINTCNEYQQRHGKKTLLFV 300  
QY 301 P 301



STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3944:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1196 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...1196  
SEQUENCE DESCRIPTION: SEQ ID NO: 3944:  
US-09-107-532A-3944  
Query Match 7.2%; Score 108.5; DB 4; Length 1196;  
Best Local Similarity 20.1%; Pred. No. 0.1;  
Matches 66; Conservative 41; Mismatches 89; Indels 133; Gaps 12;  
QY 3 GIPAEQTVGVK-----SAETADG---ALDFVNYKYLDOVPMKTFDETIKELSR 49  
DB 162 GIP-ETAGVLYKQKKAQKLFETEDNLSRVQDIHELEEQLTPLAAQSEAAKEFLR 220  
QY 50 FKQEYSQ-EASVLVGDIKVLLMDSQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKK 108  
DB 221 LKETLTQTDVSLMVAEIKTAKD-----WDNKQ 248  
QY 109 ASAQKILIRILDDGVNKLNEA---QKSLGSSQSPFNASGKLALDSOLTNDFFSEKSY 165  
DB 249 AQLAKFNL-----ELGKSESIOQESILAKQKNAQADRLIEKXQVQLDLSEKLQ 302  
QY 166 FQSQVDRIRKAYAGAAAGIVAGPFGLLIISYTAAGVIEGKLIPELNDRLKAVQNFFTSL 225  
DB 303 TEQCKD-----VLQERTKTKTQSSQY 324  
QY 226 SVTVKQAKDIDAAKULATEIAAIGETETTRFYVDYDMLSLKGAAK-----280  
DB 325 QTSIAEAK-----KVK-----HFEXLQSLMKAAAEKETEIQ 357  
QY 281 -----MINTCNEYQ---QRHGKKTLLVDP 302  
DB 358 KAEANLIKTOOLEKYQKSTKELLAEIRD 386  
RESULT 6  
US-09-134-001C-5080  
; Sequence 5080, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5080  
; LENGTH: 3696  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5080  
Query Match 7.1%; Score 106.5; DB 3; Length 3696;  
Best Local Similarity 21.5%; Pred. No. 1;  
Matches 71; Conservative 49; Mismatches 129; Indels 81; Gaps 12;  
QY 11 EVKSAETADGALDFYNYKYLDOVPMKTFDETIKELSR-----FKQEYSQ-EA 58  
DB 568 EQVNDIIPSNYTLASVNY-----NKLKERAQTVLDEETNTTFFNOKYSQTQI 616  
QY 59 SVLVGDIKVLIM-----DSQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASA 111  
DB 617 DLLLHELQTTLINRVASREINDKAQEMTDVYDSTELTTEKDT---LYVDQIENHKNEI 673  
QY 112 QKDILIRILDDGVNKLNEAOKSLIGSS-----QSFNNSAGKLLALDSOLTNDFFS 160  
DB 674 SNIDDELTDGVERVEKAGLHLESTPHVTPYKPNARQVNNRA-----DQCKTLIRN 727  
QY 161 EKSSYFQSQVDRIRK-EAYAGAAAGIVAGPFGLLIISYTAAGVIEGKLIPELNDRLKAVQ 219  
DB 728 NHEATTEQNEAIRQVEAHSSDA-----IAKIGEAETDTTVNE---ARD 768  
QY 220 NFFTSLSVTVKQAKDIDAAKULATEIAAIGETETTRFYVDYDMLSLKGAAK 279  
DB 769 NGTKLIATDVENPTKAAE---RAAVTNSANSKIKDINNNTQATLDERNDALVNRSKD 825  
QY 280 KMINTCNEYQ-----QRHGKKTLLVDP 301  
DB 826 EAIQNINTAQNDVDVTEAQNNGTNTIQVVP 855  
RESULT 7  
US-09-107-532A-5150  
; Sequence 5150, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ariniello, Pamela Denike  
 REGISTRATION NUMBER: 40,489  
 REFERENCE/DOCKET NUMBER: GTC-012  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277  
 INFORMATION FOR SEQ ID NO: 5150:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1231 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Enterococcus faecium  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (B) LOCATION 1...1231  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5150:

Query Match 7.0%; Score 106; DB 4; Length 1231;  
 Best Local Similarity 23.0%; Pred. No. 0.2;  
 Matches 77; Conservative 60; Mismatches 124; Indels 74; Gaps 17;

QY 6 AEQTVVVKSAIETADGALD-----FVKYLDQVIPWKTDE-----TIKELS 48  
 DB 269 AEEIVSGVSAQTIEQIIDYLANGRTGFLNIHLRPPFVETVQKLPHTVKTIAVL 328  
 QY 49 RFKQBSQSEASVLGDIKVLMSQ-----DKYFEATQTVYEMCGVVTQLLSAYILLF 101  
 DB 329 RSKPAGGEPILL-DVQSALYDSELPAVIGRYGLGSK-----DVTPOISA---VF 378  
 QY 102 DEYNEKASAKDILIRILDGYNKLENAQSL-LGSSQSFN-----NASGKL 148  
 DB 379 DELKQPSIRKRFITIGVDDVYQSLPESKSLDTEPTTQAKFWGFGSDGTVGANKSA 438  
 QY 149 LALDSQLTNDFSEKSYFQSQ-----VDRIK-----KEAYAGAAAGIVA--GPFGLII 194  
 DB 439 IKIIGDHTDKVAQGYFYDSSKSGGLTSHLRFGTPIRSAYLVHEDLVACHTP-AVIL 497  
 QY 195 SYSTAAGVIEGLIPELNDRKAVQNFSTLSVTVKQ--ANKDI---DAAKLKLATEIAA 249  
 DB 498 SYDLVKGLKPGGIF--LLNTLMSDQLETHLPKLKRYLAENNIRFYTNAMRLAQEVL 555  
 QY 250 IGEIKTETETTRF-----YVDYDDLMSLILKGAOK 280  
 DB 556 GRRINTAMETAPFKLADIIPFDE-VLPLKKEALK 589

RESULT 8  
 US-09-328-352-4284  
 ; Sequence 4284, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 4284  
 ; LENGTH: 815  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-4284

Query Match 6.9%; Score 104; DB 4; Length 815;  
 Best Local Similarity 22.0%; Pred. No. 0.17;  
 Matches 72; Conservative 47; Mismatches 123; Indels 86; Gaps 15;

QY 9 TVEVVKSAIETADGALDFYNKYLDQVIPWKTDETIKELSRFKQ--EYSQSEASVLVGDIK 66  
 DB 220 TAHIIQSWADKADTEIIPQITILEPGLPYEPDVEIKAESEFKEPAEKDREGVRLDLA 279  
 QY 67 VLLMDSQDKYFEATQTVYEMCGVVTQLLSAYILLFDE--YNEKASAKDILIRI----- 119  
 DB 280 LVTIDGEDARD-----FDDAVYAEKRPGGYGVVVAIVADVSH 316  
 QY 120 --LDDGVNKLNEAKSLLGSSQSFNNASGKLALDSQLTNDFSSEKSYFQSQVDRI--- 173  
 DB 317 YVRILDSALNE--EAEER--GTSVYFPHP--VLPMLPEALSNGLCSLNPH---VDRLCMV 366  
 QY 174 --RKEAYAGAAAGIVAGPFGI-----IISYSIAAGVIEG--KLIPELNDRKAVQNFSTLS 225  
 DB 367 CDLKLRSRTGRVTGYEFPVAVMHSKARLTYTQGVYFEGATDAIPKDRDIHKSNTLTF-QL 425  
 QY 226 SVTVKQANKDIDAAKLKLATIAAIGEIKTETETTRFYVDYDDL-----MLSLLKGAOK 280  
 DB 426 YQILKNLVRDRA-----MBFETIETMTFDELGGIKEILPRTENDAHK 469  
 QY 281 MINTC-----NEYQORHGKKTILLEV 300  
 DB 470 LIECNLLANVAAYALEHDIPMLYRV 497

## RESULT 9

US-09-134-001C-4820  
 ; Sequence 4820, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 4820  
 ; LENGTH: 1211  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-4820

Query Match 6.9%; Score 104; DB 3; Length 1211;  
 Best Local Similarity 18.0%; Pred. No. 0.31;  
 Matches 60; Conservative 60; Mismatches 131; Indels 82; Gaps 9;

QY 7 BOTVEVVK-----SAIETADGALDFYNKYLDQVIPWKTDETIKELSRFKQSEASV 60  
 DB 185 ESAGVLKYKKRKAESIQKLDHTEDNLNRVEDILYDLEGRVEPLKEAAIAKEAYKQLSKE 244  
 QY 61 LVGDIKVLMSQDKYFEATQTVYEMC-----GVVTQLLSAYILLFDEYNE 106  
 DB 245 MEQSDVIVTVSDIDHYTEDNQRLDERNLHLSQQAKEGQQAQINQLQKY----- 295  
 QY 107 KQASAKDILIRIIDDGVNKLNEAKSLLGSSQSFNNASGKLALDSQLTND-----FS 160  
 DB 296 -KGRQQN-----DYDIEKLN---YELVKAATYEQLSGKLNVLEERKQSQSETNARYE 345  
 QY 161 EKSSYFQSQVDRIKKEAYAGAAAGIVAGPFGIILISYSIAAGVIEGLKIPELNDRL---K 216  
 DB 346 EELNLESQIDSINKKAQNE-----KLADLNKQKQNLN 381  
 QY 217 AVQNFSTLSVTVKQANKDIDAAKLKLATEIAAIGEIKTETETTRFYVDYDDLMSLLKG 276  
 DB 382 EVOELESLLYISDSQHDKELEIKNSYVTLMSQSVVNDIRFLEHTINENEAKSRDLS 441  
 QY 277 AAKXMINTCNBYQQ-----RHGKKTILLEV 300

```

; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-1660

```

|                       |        |                    |  |                          |                    |
|-----------------------|--------|--------------------|--|--------------------------|--------------------|
| Query Match           | 6.8%;  | Score 103;         | DB 4;  | Length 885;              |                    |
| Best Local Similarity | 18.2%; | Midat.No. 0.24;    |  |                          |                    |
| Matches               | 53;    | Conservative       | 52;  | Mismatches 111;          | Indels 76; Gaps 8; |
|                       |        |                    |  |                          |                    |
| Qy                    | 42     | ETIKELSRFKQOEYSQEA | SVLIGDIKVLMLDMSQDKYPEATQTVYWC                | -----                    | 87                 |
|                       |        |                    |  |                          |                    |
| Db                    | 2      | EPLKEEAIAKEYKQ     | SKEMEQSDVIWTVTSDIDHYTENDQRLDERLNLKSKQAKEGQQ  | 61                       |                    |
|                       |        |                    |  |                          |                    |
| Qy                    | 88     | GVVTKQLLSAYILL     | FDENYNEKSAQAKDILIRILDDGVNKLNEAQKSLIGSSQSFNNA | 147                      |                    |
|                       |        |                    |  |                          |                    |
| Db                    | 62     | AQINQLLOKY         | -----KGRKQON-----DYDIEKLN--YELVKATENYQLSGK   | 102                      |                    |
|                       |        |                    |  |                          |                    |
| Qy                    | 148    | LIALDSQLTND        | -----FSBKSSYFOSQVDRI                         | NRKEAYAGAAAGIVAGPFLGIISY | 201                |
|                       |        |                    |  |                          |                    |
| Db                    | 103    | LVNLVEERKKNQ       | SETNARYEEELONLESQIDSINKEA                    | QONE-----                | 142                |
|                       |        |                    |  |                          |                    |
| Qy                    | 202    | VIEGKLIPELNDRL     | -----KAVQNFFTSLVTVVKQANKDIAAKL               | KLATEIAAIGIKET           | 257                |
|                       |        |                    |  |                          |                    |
| Db                    | 143    | -----KLIAELNKKOK   | LNKEVQELLESLLIYSDSQHDEKLEEKNSYTYLLMSQSDV     | NNDI                     | 198                |
|                       |        |                    |  |                          |                    |
| Qy                    | 258    | ETTFREYVDYD        | DLMLSLKGAAGKMTNC                             | EYQQ-----RHGKTYLLEV      | 300                |
|                       |        |                    |  |                          |                    |
| Db                    | 199    | RLFLEHTINENAK      | KSRILSDSRIVEAFNLGDKIQNNITQTK                 | QYQSSKSKMEKV             | 250                |
|                       |        |                    |  |                          |                    |

```

RESULT 12
US-09-914-259-66
; Sequence 66, Application US/09914259
; Patent No. 6495356
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hymen, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-914-259-66

```

Query Match 6.8%; Score 103; DB 4; Length 961;  
Best Local Similarity 22.0%; Pred. No. 0.27;  
Matches 70; Conservative 53; Mismatches 129; Indels 66; Gaps 14

|    |     |   |                      |      |
|----|-----|---|----------------------|------|
| QY | 11  | EVKSAIETADGALDPFNKYL  | -----DQVPMKTPDE----- | 6998 |
| Db | 639 | EVKTLIEHDSIVTHYKNMIREQDLLEELKQIISTLKQNEQLQTAVTQQVSIQOQH         |                      | 6998 |
| QY | 52  | QEYSQERASVLVGDIKVLL--WDSQDK--YFEATQTVVWCGVVVTTOLLGAYILLFPEYNEKK |                      | 108  |
| Db | 699 | DOYNL-----LKVLQKDSHQGPYTDGAQ-----MNGVQPEISR--LREEIEELK          |                      | 743  |
| QY | 109 | AS-----AQKDILIRIIDGVNK--LNEAQSLLGSSOFNNSAGKLLALDSQLTND          |                      | 158  |
| Db | 744 | SNRELLOSQAEXDSLLENKSSQLSPGTNEQSSATAGDSQIAELKQELATLKSQ--NS       |                      | 802  |
| QY | 159 | FSEKSSYFQSQVDRIRKEAVAGAAAGIVAGPGLIISIAANGVEGL-----IPELN         |                      | 212  |
| Db | 803 | OSVEITTKLOTEKQELLOKTEFAKAPVPGESETVIATKTD--VEGRSALLQETKELK       |                      | 866  |
| QY | 213 | DLKAVQNFFTSLSVTVKQANKIDAAUKLATEIAAIGEIKTETTRTFVVDVDDLMS         |                      | 277  |
| Db | 861 | NEIKALGEERTAIKEQDSSNSTI-----AILONEKNKLEVDITDSKEQDOLLV--         |                      | 91   |

```
db      :   :    : |::|  
442 RLVEAFNLKDIIQQNITQTQKEYSSKKSMEKV 474  
  
RESULT 10  
US-09-134-000C-3594  
; Sequence 3594, Application US/09134000C  
; Patent No., 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032798-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3594  
; LENGTH: 284  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
; US-09-134-000C-3594
```

|                       |        |   |  |                    |  |
|-----------------------|--------|---|--|--------------------|--|
| Query Match           | 6.8%;  | Score 103;  | DB 4;  | Length 284;        |  |
| Best Local Similarity | 21.0%; | Pred. No. 0.039;  |  |                    |  |
| Matches               | 67;    | Mismatches 46;  | Indels 100;  | Gaps 13            |  |
| Conservative          |        |   |  |                    |  |
| QY                    | 16     | A I E T A D G A L D F Y N K Y L D O V I P W K Y F D E T I K E L   | -----SRFQ E Y S Q R A S V L V G D I K V L  | --- 68             |  |
| Db                    | 6      | A V K V G G E G A N F L E K K E K   | -----Y D Y T I K N F D D A T G L Y K A L E N G E A D A I V D D Y P V L G V A           | 59                 |  |
| QY                    | 69     | -----L M D S Q D  | -----K Y F E A T Q T V Y E W C V V T Q L L S A V I L L F D                             | 102                |  |
| Db                    | 60     | V K N G Q K L Q L V G D K E T G S S Y G F A V K K G O N P E L I K F N A G L K N L K D N G Y F D K I L N Y I L A T G D | -----F N N A S G K L J A L D S Q L T   | 119                |  |
| QY                    | 103    | E Y N E K K A S A Q K D I L R I L D D G V N K L N E A Q K S I L G S S O S   | -----F N N A S G K L J A L D S Q L T   | 156                |  |
| Db                    | 120    | E T W T Q A G E Q   | -----M K K I T P K K E X V I A S D T F A P P E F Q N A Q D Y V G I D V D V L           | 168                |  |
| QY                    | 157    | N D P S E K S S Y   | -----F O S Q V D R I R K E A Y A G A G I V A G P F G L I I S Y S I A A G V I E G K L I | 208                |  |
| Db                    | 169    | K R A E L Q G F T V E F K F I G F S S A Q V A E   | -----S Q O A D G M V A G   | ---M 205           |  |
| QY                    | 209    | P E L N D R L K A   | -----V Q N F F T S L S V T K V K O A N   | -----K D I D A A K | ---L K L A T E I A - A I G E I K T 255 |
| Db                    | 265    | -----   | -----  | -----              | -----                                  |

QY            256 ETETTRFYVDYDMLLSLL 274  
             : : | : | : | :  
Db            266 KYDISIKYLDTTDALYSAL 284

RESULT 11  
US-09-710-279-1660  
; Sequence 1660, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1660  
; LENGTH: 885  
; TYPE: PR1  
; ORGANISM: Artificial Sequence  
; FEATURE:





us-09-993-292b-24.ra1

Wed Feb 2 11:26:40 2005

Db 1313 LSSQLQDQELQETROK-----LNVTSLKQLQLEBERNSLOQDLDEMEAK 1359  
Qy 219 QNF---FTSLSVTVKQANKDIDAAKLKLAIEAIGIKTETETTRFYVDYDDLMLSLK 275  
Db 1360 QNLERHISTNLQISDSKKLQ---DPASTVEALEBK-----KRFQKIEINLTQOYEE 1410  
Qy 276 GAA--KKMINTCNEYQQ 290  
Db 1411 KAAAYDKLEKYNRLQQ 1427

RESULT 16  
US-09-543-681A-8019  
; Sequence 8019, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 8019  
; LENGTH: 1129  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-8019

Query Match 6.7%; Score 101; DB 4; Length 1129;  
Best Local Similarity 23.8%; Pred. No. 0.56;  
Matches 68; Conservative 48; Mismatches 128; Indels 42; Gaps 13;  
Qy 4 IFAEQTVVVKSAIEATGADGALDFYKYLDOVIPWKTFTETIKELGRFKQOYSEASVLVG 63  
Db 435 IADWLEENQTSINKAGSLDRAETIGDKA-----RNASQAERLGLLEI-KEADSMG 487  
Qy 64 DIKVLMDSQKYFEATQTVYEWGVVTL--LSAYILLFDEYNEKKAQKIDILRIID 121  
Db 488 K-KSLMDINKY---QKIINERSAVEVELNSLAKLNKTQTRQATLORK--LQRIN 541  
Qy 122 DGVNKINEAKS---LGSOSFNNASCKLLALDSQTLNDESEKSYFQSQVDRIKEAVA 179  
Db 542 DAENKLPALQSVLDNPKFNEHRLTRTANSLTRHDRIQSAINRMTPLEREELDA 601  
Qy 180 GA---AAGIVAGPGLIISYSIAAGVIEGK-----LIPE--LNDRLKAVQNF-- 222  
Db 602 AADIVNKIICAPSGIYVSELPGLVKRAGFTKDRTLNIPDERIKDYLEDVNVYVWYNY 661  
Qy 223 ----TSLSVTVKQANKDIDAAKLKLAIEAIGIKTETETTR 261  
Db 662 IRQVAPEIELTAKGRVDMQ-NQIKAITVEYNQLADIATTPKRSR 706

RESULT 17  
US-09-447-497-15  
; Sequence 15, Application US/09447497  
; Patent No. 6773911  
; GENERAL INFORMATION:  
; APPLICANT: Penninger, Josef M.  
; APPLICANT: Kroemer, Guido P.  
; APPLICANT: Siderovski, David P.  
; APPLICANT: Zamzami, Naoufal  
; APPLICANT: Susin, Santos A.  
; APPLICANT: Snow, Bryan E.L.  
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR  
; FILE REFERENCE: 01017/36780  
; CURRENT APPLICATION NUMBER: US/09/447,497  
; CURRENT FILING DATE: 1999-11-23  
; EARLIER APPLICATION NUMBER: 60/109,595

; EARLIER FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 526  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human apoptosis-inducing factor (half isoform #2);  
; OTHER INFORMATION: half-exon-skip-Gold; mature polypeptide  
US-09-447-497-15

Query Match 6.7%; Score 100.5; DB 4; Length 526;  
Best Local Similarity 22.6%; Pred. No. 0.19;  
Matches 68; Conservative 50; Mismatches 94; Indels 89; Gaps 16;  
Qy 17 IETADGALDFYKYLDOVIPWKTFTETIKELSRFKQOYSEAS----- 59  
Db 74 IVSEDPPELPMRPPLSKEL-WFSDDPNVTKTLRFQWNGKERSIYFQPPSFVVSADLPH 132  
Qy 60 -----VLVGDIDKILLMDSQD---KYFEATQTVYEWGVVTL---QLLSAYILLFDEVN 105  
Db 133 IENGVAVLTKG-KVQVLDVRDNMVKLNDGSOITVEKCLIAATGTPRSLAI- 183  
Qy 106 EKASQAQ---KDILIRILD--GVNKLNEAQKSL-----LGSOSFNNASG-KILALD 152  
Db 184 -DRAGAEVKSRTTLFRKIGDFRSLEKISREVKSITTIIGGFLGSELAC--ALGRKARALG 240  
Qy 153 SOLTNDPSEKSS-----YFQS-QVDRIRKAYAGAAAGIVAGPGLIISYSIAAGVIE 204  
Db 241 TEVIQLPPEKGGKILPELSNWTKEVRE-----GVKMPNAIVQS---VGVS 289  
Qy 205 GKLIPELNDRLKAVQNFFTSLSVTVKQANKDIDAAKLKLAIEAIGIKTETETTRFYV 264  
Db 290 GKLLIKLKDGRKV-----ETDHVAAGVLEPNVELAKTGGLLEIDSDFGGRV 336  
Qy 265 D 265  
Db 337 N 337

RESULT 18  
US-09-447-497-14  
; Sequence 14, Application US/09447497  
; Patent No. 6773911  
; GENERAL INFORMATION:  
; APPLICANT: Penninger, Josef M.  
; APPLICANT: Kroemer, Guido P.  
; APPLICANT: Siderovski, David P.  
; APPLICANT: Zamzami, Naoufal  
; APPLICANT: Susin, Santos A.  
; APPLICANT: Snow, Bryan E.L.  
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR  
; FILE REFERENCE: 01017/36780  
; CURRENT APPLICATION NUMBER: US/09/447,497  
; CURRENT FILING DATE: 1999-11-23  
; EARLIER APPLICATION NUMBER: 60/109,595  
; EARLIER FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 553  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-447-497-14

Query Match 6.7%; Score 100.5; DB 4; Length 553;  
Best Local Similarity 22.6%; Pred. No. 0.2;  
Matches 68; Conservative 50; Mismatches 94; Indels 89; Gaps 16;  
Qy 17 IETADGALDFYKYLDOVIPWKTFTETIKELSRFKQOYSEAS----- 59  
Db 101 IVSEDPPELPMRPPLSKEL-WFSDDPNVTKTLRFQWNGKERSIYFQPPSFVVSADLPH 159

Qy 60 -----VLVGDIKVLLMDSOD---KYFEATQTVVEMCGVVT-----QLLSAYILLFDEYN 105  
Db 160 IENGGAVALTGK-KVQVLDVRDNMVKLNDGSGQITVEKCLATGTPRSLSAI-----210  
Qy 106 EKKASQAQ---KDILIRILDD--GVNKLNEAOKSL-----LGSSQSFNNASG-KLLALD 152  
Db 211 -DRAGAEVKSRTTLFRKIGDFRSLEKISREVKSITIIIGGFLGSELAC--ALGRKARALG 267  
Qy 153 SOLTNDSEKSS-----YFQS-QVDRIRKEAYAGAAAGIVAGPFGLLIISYSAAGVIE 204  
Db 268 TEVIQLPPEKGMGKILPEYLSNWTMEKVRRE-----GVKMPNAIVQS-----VGVSS 316  
Qy 205 GKLIPELNDRLKAVQNFTSLSVTVKQAKDIDAOKLKLATEIAAIGETETETTRFYV 264  
Db 317 GKLLIKLKDGRKV-----ETDHIHVAAGLEPNVELAKTGGLDSDFGGFRV 363  
Qy 265 D 265  
Db 364 N 364

## RESULT 19

US-09-538-092-807  
; Sequence 807, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538.092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuratSeqFormer Version 0.9  
; SEQ ID NO 807  
; LENGTH: 566  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (0)..(0)  
; OTHER INFORMATION: Polypeptide Accession Number YPR070W  
US-09-538-092-807

Query Match 6.7%; Score 100.5; DB 4; Length 566;  
Best Local Similarity 25.0%; Pred. No. 0.21;  
Matches 56; Conservative 29; Mismatches 92; Indels 47; Gaps 11;  
Qy 15 SAITADGALDFYK-----LDQVTPWKTDFETI-KELSRFKO-----52  
Db 6 SYVETLDSMIELFDYKPGSITLENITRLCQTGLGSEFTELSNELSRSLTASKIIVIDV 65  
Qy 53 EYSEASVLVGDIKVLLMDSODK--YFEATQTVVEMCGVVTQLLSAYILLFDEYNEKKAS 110  
Db 66 DYNKKQD-RIDQVKLVLASNFEDYFNQRGEHEKSNILNSLTKYPDLLKAFHNNLKL 124  
Qy 111 AQDKILIRILDDGVNKLN-EAOKSLGSSQSFNNASGKLLALDSQLTNDSEKSSY----165  
Db 125 YLLDAYSHIESDTSIHNNGSSDKSLDSSNFSNN-QGKL-----DLPKYFTLSHYLRQC 178  
Qy 166 FQSQVDRIRKEAYAGAAAGIVAGPFGLLIISYSAAGVIEGKLIP 209  
Db 179 FQDNCCDFKVRTNLNDKFGI-----YILTQG-INGKEVP 211

## RESULT 20

US-09-447-497-12  
; Sequence 12, Application US/09447497

; Patent No. 6773911  
; GENERAL INFORMATION:  
; APPLICANT: Penninger, Josef M.  
; APPLICANT: Kroemer, Guido P.  
; APPLICANT: Siderovski, David P.  
; APPLICANT: Zamzami, Naoufal  
; APPLICANT: Susin, Santos A.  
; APPLICANT: Snow, Bryan E.L.  
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR  
; FILE REFERENCE: 01017/36780  
; CURRENT APPLICATION NUMBER: US/09/447.497  
; EARLIER FILING DATE: 1999-11-23  
; EARLIER FILING DATE: 1999-11-23  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 609  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human apoptosis-inducing factor (half isoform #1  
; OTHER INFORMATION: ); half-alt-exon-Gold; mature polypeptide  
US-09-447-497-12  
Query Match 6.7%; Score 100.5; DB 4; Length 609;  
Best Local Similarity 22.6%; Pred. No. 0.24;  
Matches 68; Conservative 50; Mismatches 94; Indels 89; Gaps 16;

Qy 17 IETADGALDFYKYLDOVTPWKTDFETIKLSRFPKQYSQAS-----59  
Db 157 IVSEDPPELPMRPPLSKEL-WFSDDPNVTITLRFKQWNGKRSITVFQPPSYVSAQDLPH 215  
Qy 60 -----VLVGDIKVLLMDSOD---KYFEATQTVVEMCGVVT-----QLLSAYILLFDEYN 105  
Db 216 IENGGAVALTGK-KVQVLDVRDNMVKLNDGSGQITVEKCLATGTPRSLSAI-----266  
Qy 106 EKKASQAQ---KDILIRILDD--GVNKLNEAOKSL-----LGSSQSFNNASG-KLLALD 152  
Db 267 -DRAGAEVKSRTTLFRKIGDFRSLEKISREVKSITIIIGGFLGSELAC--ALGRKARALG 323  
Qy 153 SOLTNDSEKSS-----YFQS-QVDRIRKEAYAGAAAGIVAGPFGLLIISYSAAGVIE 204  
Db 324 TEVIQLPPEKGMGKILPEYLSNWTMEKVRRE-----GVKMPNAIVQS-----VGVSS 372  
Qy 205 GKLIPELNDRLKAVQNFTSLSVTVKQAKDIDAOKLKLATEIAAIGETETETTRFYV 264  
Db 373 GKLLIKLKDGRKV-----ETDHIHVAAGLEPNVELAKTGGLDSDFGGFRV 419  
Qy 265 D 265  
Db 420 N 420

## RESULT 21

US-09-447-497-9  
; Sequence 9, Application US/09447497  
; Patent No. 6773911  
; GENERAL INFORMATION:  
; APPLICANT: Penninger, Josef M.  
; APPLICANT: Kroemer, Guido P.  
; APPLICANT: Siderovski, David P.  
; APPLICANT: Zamzami, Naoufal  
; APPLICANT: Susin, Santos A.  
; APPLICANT: Snow, Bryan E.L.  
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR  
; FILE REFERENCE: 01017/36780  
; CURRENT APPLICATION NUMBER: US/09/447.497  
; CURRENT FILING DATE: 1999-11-23  
; EARLIER FILING DATE: 1999-11-23  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0

|   |     |   |     |
|---|-----|---|-----|
| Db  | 230 | IENGGAVALTGK-KVQLDVRDNNMVKLNDGSGQITTEKLIATGCTPRSLSAI-----     | 280 |
| QY  | 106 | EKASAQ---KDILIRILDD--GVNKLNEAQKSL-----LGSSQSFNNASG-KLIALD     | 152 |
| Db  | 281 | ---DRAGAEVKSRTTLFRKIGDPRSLEKISREVKSITIIIGGFLGSELAC--ALGKARALG | 337 |
| QY  | 153 | SOLTNDPSEKSS-----YFQS-QVDRIRKEAYAGAAAGIVAGPFGLIISYIAAGVIE     | 204 |
| Db  | 338 | TEVIQLPFPEKGNMGKILPEYLSNWTMEKVRE-----GVKVMFNAIVOS---VGVSS     | 386 |
| QY  | 205 | GKLIPELNDRLKAVQNFTSLSVTVKQAKDIDAAKLAATEIAAIGETETETTRFYV       | 264 |
| Db  | 387 | GKLLIKLKDGRKV-----ETDHIVAAGLEPNVELAKTGGLGLEIDSDFGGFRV         | 433 |
| QY  | 265 | D 265   |     |
| Db  | 434 | N 434   |     |
| RESULT 23   |     |   |     |
| US-09-447-497-8   |     |   |     |
| ; Sequence 8, Application US/09447497                           |     |   |     |
| ; Patent No. 6773911  |     |   |     |
| ; GENERAL INFORMATION:  |     |   |     |
| ; APPLICANT: Penninger, Josef M.                                |     |   |     |
| ; APPLICANT: Kroemer, Guido P.                                  |     |   |     |
| ; APPLICANT: Siderovski, David P.                               |     |   |     |
| ; APPLICANT: Zamzami, Naoufal                                   |     |   |     |
| ; APPLICANT: Susin, Santos A.                                   |     |   |     |
| ; APPLICANT: Snow, Bryan E.L.                                   |     |   |     |
| ; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR                 |     |   |     |
| ; FILE REFERENCE: 01017/36780                                   |     |   |     |
| ; CURRENT APPLICATION NUMBER: US/09/447,497                     |     |   |     |
| ; EARLIER FILING DATE: 1999-11-23                               |     |   |     |
| ; EARLIER FILING DATE: 1998-11-23                               |     |   |     |
| ; NUMBER OF SEQ ID NOS: 25                                      |     |   |     |
| ; SOFTWARE: Patent in Ver. 2.0                                  |     |   |     |
| ; SEQ ID NO 8   |     |   |     |
| ; LENGTH: 627   |     |   |     |
| ; TYPE: PRT   |     |   |     |
| ; ORGANISM: Homo sapiens  |     |   |     |
| US-09-447-497-8   |     |   |     |
| Query Match 6.7%; Score 100.5; DB 4; Length 627;                |     |   |     |
| Best Local Similarity 22.6%; Pred. No. 0.25;                    |     |   |     |
| Matches 68; Conservative 50; Mismatches 94; Indels 89; Gaps 16; |     |   |     |
| QY  | 17  | IETADGALDFYKYLQDVIPWKTFTETIKELSRFKQYSQEAS-----                | 59  |
| Db  | 175 | IVSEDPPELPYMRPPLSKEL-WFSDDPNVTKTLRFKQWNGKRSIYFQPPSFYVSAQDLPH  | 233 |
| QY  | 60  | -----VLVGDIKVLLMDSQD---KYFEATQTVYEWCGVVT---QLLSAYILLFDEYN     | 105 |
| Db  | 234 | IENGGAVALTGK-KVQLDVRDNNMVKLNDGSGQITTEKLIATGCTPRSLSAI-----     | 284 |
| QY  | 106 | EKASAQ---KDILIRILDD--GVNKLNEAQKSL-----LGSSQSFNNASG-KLIALD     | 152 |
| Db  | 285 | ---DRAGAEVKSRTTLFRKIGDPRSLEKISREVKSITIIIGGFLGSELAC--ALGKARALG | 341 |
| QY  | 153 | SOLTNDPSEKSS-----YFQS-QVDRIRKEAYAGAAAGIVAGPFGLIISYIAAGVIE     | 204 |
| Db  | 342 | TEVIQLPFPEKGNMGKILPEYLSNWTMEKVRE-----GVKVMFNAIVOS---VGVSS     | 390 |
| QY  | 205 | GKLIPELNDRLKAVQNFTSLSVTVKQAKDIDAAKLAATEIAAIGETETETTRFYV       | 264 |
| Db  | 391 | GKLLIKLKDGRKV-----ETDHIVAAGLEPNVELAKTGGLGLEIDSDFGGFRV         | 437 |
| QY  | 265 | D 265   |     |
| Db  | 438 | N 438   |     |

|   |     |   |     |
|---|-----|---|-----|
| Db  | 230 | IENGGAVALTGK-KVQLDVRDNNMVKLNDGSGQITTEKLIATGCTPRSLSAI-----     | 280 |
| QY  | 106 | EKASAQ---KDILIRILDD--GVNKLNEAQKSL-----LGSSQSFNNASG-KLIALD     | 152 |
| Db  | 281 | ---DRAGAEVKSRTTLFRKIGDPRSLEKISREVKSITIIIGGFLGSELAC--ALGKARALG | 337 |
| QY  | 153 | SOLTNDPSEKSS-----YFQS-QVDRIRKEAYAGAAAGIVAGPFGLIISYIAAGVIE     | 204 |
| Db  | 338 | TEVIQLPFPEKGNMGKILPEYLSNWTMEKVRE-----GVKVMFNAIVOS---VGVSS     | 386 |
| QY  | 205 | GKLIPELNDRLKAVQNFTSLSVTVKQAKDIDAAKLAATEIAAIGETETETTRFYV       | 264 |
| Db  | 387 | GKLLIKLKDGRKV-----ETDHIVAAGLEPNVELAKTGGLGLEIDSDFGGFRV         | 433 |
| QY  | 265 | D 265   |     |
| Db  | 434 | N 434   |     |
| RESULT 23   |     |   |     |
| US-09-447-497-8   |     |   |     |
| ; Sequence 8, Application US/09447497                           |     |   |     |
| ; Patent No. 6773911  |     |   |     |
| ; GENERAL INFORMATION:  |     |   |     |
| ; APPLICANT: Penninger, Josef M.                                |     |   |     |
| ; APPLICANT: Kroemer, Guido P.                                  |     |   |     |
| ; APPLICANT: Siderovski, David P.                               |     |   |     |
| ; APPLICANT: Zamzami, Naoufal                                   |     |   |     |
| ; APPLICANT: Susin, Santos A.                                   |     |   |     |
| ; APPLICANT: Snow, Bryan E.L.                                   |     |   |     |
| ; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR                 |     |   |     |
| ; FILE REFERENCE: 01017/36780                                   |     |   |     |
| ; CURRENT APPLICATION NUMBER: US/09/447,497                     |     |   |     |
| ; EARLIER FILING DATE: 1999-11-23                               |     |   |     |
| ; EARLIER FILING DATE: 1998-11-23                               |     |   |     |
| ; NUMBER OF SEQ ID NOS: 25                                      |     |   |     |
| ; SOFTWARE: Patent in Ver. 2.0                                  |     |   |     |
| ; SEQ ID NO 8   |     |   |     |
| ; LENGTH: 627   |     |   |     |
| ; TYPE: PRT   |     |   |     |
| ; ORGANISM: Homo sapiens  |     |   |     |
| US-09-447-497-8   |     |   |     |
| Query Match 6.7%; Score 100.5; DB 4; Length 627;                |     |   |     |
| Best Local Similarity 22.6%; Pred. No. 0.25;                    |     |   |     |
| Matches 68; Conservative 50; Mismatches 94; Indels 89; Gaps 16; |     |   |     |
| QY  | 17  | IETADGALDFYKYLQDVIPWKTFTETIKELSRFKQYSQEAS-----                | 59  |
| Db  | 175 | IVSEDPPELPYMRPPLSKEL-WFSDDPNVTKTLRFKQWNGKRSIYFQPPSFYVSAQDLPH  | 233 |
| QY  | 60  | -----VLVGDIKVLLMDSQD---KYFEATQTVYEWCGVVT---QLLSAYILLFDEYN     | 105 |
| Db  | 234 | IENGGAVALTGK-KVQLDVRDNNMVKLNDGSGQITTEKLIATGCTPRSLSAI-----     | 284 |
| QY  | 106 | EKASAQ---KDILIRILDD--GVNKLNEAQKSL-----LGSSQSFNNASG-KLIALD     | 152 |
| Db  | 285 | ---DRAGAEVKSRTTLFRKIGDPRSLEKISREVKSITIIIGGFLGSELAC--ALGKARALG | 341 |
| QY  | 153 | SOLTNDPSEKSS-----YFQS-QVDRIRKEAYAGAAAGIVAGPFGLIISYIAAGVIE     | 204 |
| Db  | 342 | TEVIQLPFPEKGNMGKILPEYLSNWTMEKVRE-----GVKVMFNAIVOS---VGVSS     | 390 |
| QY  | 205 | GKLIPELNDRLKAVQNFTSLSVTVKQAKDIDAAKLAATEIAAIGETETETTRFYV       | 264 |
| Db  | 391 | GKLLIKLKDGRKV-----ETDHIVAAGLEPNVELAKTGGLGLEIDSDFGGFRV         | 437 |
| QY  | 265 | D 265   |     |
| Db  | 438 | N 438   |     |

RESULT 24  
 US-08-938-105-3  
 ; Sequence 3, Application US/08938105  
 ; Patent No. 6353151  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weinwand, Leslie A.  
 ; APPLICANT: Vikstrom, Karen L.  
 ; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sheridan Ross P.C.  
 ; STREET: 1700 Lincoln St., Suite 3500  
 ; CITY: Denver  
 ; STATE: CO  
 ; COUNTRY: U.S.A.  
 ; ZIP: 80203  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/938.105  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Crook, Wannell M.  
 ; REGISTRATION NUMBER: 31,071  
 ; REFERENCE/DOCKET NUMBER: 3595-4  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (303) 863-9700  
 ; TELEFAX: (303) 863-0223  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1886 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-938-105-3

Query Match 6.6%; Score 100; DB 3; Length 1886;  
 Best Local Similarity 21.3%; Pred. No. 1.6; Indels 88; Gaps 12;  
 Matches 71; Conservative 44; Mismatches 131; Indels 88; Gaps 12;  
 Qy 5 PAEQTVVVKSAIETADGALDFYNYKLDQVTPWKTFDTIKELSRFKQEQYSQEQASVLVGD 64  
 Db 1120 FQWRDLLEATLQHEATAAALRKKHADSV---AELGEQIDNLRQVKLEKEKS-----E 1172  
 Qy 65 IKVLLMSQDKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKASAKQDILIRLDDGV 124  
 Db 1173 FQWRDLLEATLQHEATAAALRKKHADSV---AELGEQIDNLRQVKLEKEKS-----E 1205  
 Qy 125 N-----KLINEAQSLGSSQSFNNASGKLLALDSQLTNDPSEKS-----SYFQSQ 169  
 Db 1206 NEYRVKLEEAQRSL-----NDFTQRAKLQTEGELARQLEKEALIWQTRKLSYQQM 1261  
 Qy 170 VDIRKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPE-----LNDR 214  
 Db 1262 EDLKRQLEEEGKAKNALA-----HALQSAHDCDLLREQYEEEMEAELQRLVLSKA 1313  
 Qy 215 LKAVONFTSLSVTVKQANKDIDAQKLKATE-----IAAIGIKETETTT--RFYVD 265  
 Db 1314 NSEVAQWRKTYETDAIQRTBLEEAKKLAQRLQDAEAEVAVNAKCSLSLEKTKHRLQNE 1373  
 Qy 266 YDDLMLSLKGAACKMINTCNEYQORHGKKTLLR 299  
 Db 1374 IEDLMVDVVERSNA---AALDKKQORNFDKILAE 1404

RESULT 25  
 US-09-310-187A-1  
 ; Sequence 1, Application US/09310187A  
 ; Patent No. 6358751

; GENERAL INFORMATION:  
 ; APPLICANT: Benichou, Gilles  
 ; APPLICANT: Fedoseyeva, Eugenia  
 ; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac  
 ; TITLE OF INVENTION: Graft Rejection  
 ; FILE REFERENCE: UCSF-090  
 ; CURRENT APPLICATION NUMBER: US/09/310.187A  
 ; CURRENT FILING DATE: 1999-05-12  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 1939  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-310-187A-1  
 Query Match 6.6%; Score 100; DB 3; Length 1939;  
 Best Local Similarity 21.3%; Pred. No. 1.7; Indels 88; Gaps 12;  
 Matches 71; Conservative 44; Mismatches 131; Indels 88; Gaps 12;  
 Qy 5 PAEQTVVVKSAIETADGALDFYNYKLDQVTPWKTFDTIKELSRFKQEQYSQEQASVLVGD 64  
 Db 1173 FQWRDLLEATLQHEATAAALRKKHADSV---AELGEQIDNLRQVKLEKEKS-----E 1225  
 Qy 65 IKVLLMSQDKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKASAKQDILIRLDDGV 124  
 Db 1226 FKLLEDD-----VTSNMEQII-----KAKANLEKVSRTLEDQA 1258  
 Qy 125 N-----KLINEAQSLGSSQSFNNASGKLLALDSQLTNDPSEKS-----SYFQSQ 169  
 Db 1259 NEYRVKLEEAQRSL-----NDFTQRAKLQTEGELARQLEKEALISQTRKLSYQQM 1314  
 Qy 170 VDIRKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPE-----LNDR 214  
 Db 1315 EDLKRQLEEEGKAKNALA-----HALQSAHDCDLLREQYEEEMEAELQRLVLSKA 1366  
 Qy 215 LKAVONFTSLSVTVKQANKDIDAQKLKATE-----IAAIGIKETETTT--RFYVD 265  
 Db 1367 NSEVAQWRKTYETDAIQRTBLEEAKKLAQRLQDAEAEVAVNAKCSLSLEKTKHRLQNE 1426  
 Qy 266 YDDLMLSLKGAACKMINTCNEYQORHGKKTLLR 299  
 Db 1427 IEDLMVDVVERSNA---AALDKKQORNFDKILAE 1457

RESULT 26  
 US-09-538-092-917  
 ; Sequence 917, Application US/09538092  
 ; Patent No. 6753314  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Giot, Loic  
 ; APPLICANT: Mansfield, Traci A.  
 ; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
 ; FILE REFERENCE: 15966-542  
 ; CURRENT APPLICATION NUMBER: US/09/538.092  
 ; CURRENT FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: 60/127,352  
 ; PRIOR FILING DATE: 1999-04-01  
 ; PRIOR APPLICATION NUMBER: 60/178,965  
 ; PRIOR FILING DATE: 2000-02-01  
 ; NUMBER OF SEQ ID NOS: 1387  
 ; SOFTWARE: CuraPatSeqFormatter Version 0.9  
 ; SEQ ID NO 917  
 ; LENGTH: 1939  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (0)...(0)  
 ; OTHER INFORMATION: Polypeptide Accession Number P13533  
 ; US-09-538-092-917

Query Match 6.6%; Score 100; DB 4; Length 1939;

|    |                       |   |   |                 |             |         |
|----|-----------------------|---|---|-----------------|-------------|---------|
|    | Query Match           | 6.6%;   | Score 99.5;                                   | DB 4;           | Length 924; |         |
|    | Best Local Similarity | 17.6%;  | Pred. No. 0.58;                               |                 |             |         |
|    | Matches 50;           | Conservative  | 60;   | Mismatches 103; | Indels 71;  | Gaps 8; |
| Qy | 38                    | KTFDETTKELSR----  | PKQEYSQBASVLVGDGIKVLLMSDSDKYFEATQTVVWCGVVVTQL | 93              |             |         |
| Db | 235                   | KKYVDISIKDLKODFLAFKQEAAGIINLV--NLDLQSLSELKOKIDLIENTKND---     | QIKEL   | 290             |             |         |
| Qy | 94                    | LSAVITLFDPEYNEKKASQAQDKILIRILDGDNVKLNENAKSILSGSSQSFNNA-----   | 144   |                 |             |         |
| Db | 291                   | VSEHELIQIEKISKDLTEKFKLIVETQLLSKHESHVQQFTKELIAESKSQQVVEELEWKMT | 350   |                 |             |         |
| Qy | 145                   | -----SKGLALDSOLTWNDFSEKSYFSQSQVDRIRKEAYAGAAAGIVAGPFGIIITSYA   | 199   |                 |             |         |
| Db | 351                   | IHKADSAARILETQTSRANAKESDXYKLTD-----                           | 382   |                 |             |         |
| Qy | 200                   | AGVIEGKLPIELNDRLKAVQNFFTSLSVTTVQANKDIDAAKLIKATEIAAIGETKTET    | 259   |                 |             |         |
| Db | 383                   | ----TSEIVNDLSQIETLKANLNKL-----EEREIQNKKDQVUSELKELKELVE---     | 430   |                 |             |         |
| Qy | 260                   | TRFYDYDDLMLSLKGAAKGMINTCNEYQQORHG-----KKTILE                  | 299   |                 |             |         |

```

RESULT 29
US-09-723-820-8
; Sequence 8, Application US/09723820
; Patent NO. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-723-820-8

```

Query Match 6.6%; Score 99.5; DB 4; Length 1066;  
 Best Local Similarity 22.5%; Pred. No. 0.73;  
 Matches 65; Conservative 58; Mismatches 109; Indels 57; Gaps 14;

|    |     |  |     |
|----|-----|--|-----|
| Qy | 24  | LDFYNYKLDQVTPWKTFDETTELKELSRPQEQYSQBASVLVGDIKVLLMDSODKYFEATQTV | 83  |
| Db | 159 | LELYNEELCDL---STDDTK--IRIFDDSTKKGSVIIQGLEBEPVHSDVDYVKYLEKG     | 213 |
| Qy | 84  | YEWCGVVTTQLLSAYILLFDEYNKKASAQDKILIRILDG-----VNKNKAQSKLL        | 135 |
| Db | 214 | KERRKTATTLMNA-----QSSRSHVTFSIVVHIENGIEGEDMLKIGKLNLV--DLA       | 263 |
| Qy | 136 | GS---SOSTN-----NASGKLALDSQLTN--DFSEKSYSQSQVDRIRKEAYAG          | 180 |
| Db | 264 | GSENVSKAGNEKGIRVRETVMINOSLTGLRVITALVDRAHPYPVRESKLTRELLQESLG    | 323 |
| Qy | 181 | AA-----AGIVAGPFGL-----ITSYSTAAGVIEGKLIPELNRL--KAVQNFTSLSVT     | 228 |
| Db | 324 | RKTSIATIISFGHKDIBETJLSTLEYAHRAKNIQNK--PEVNQKLPKCTVLYKEYTE---   | 378 |
| Qy | 229 | VQOANKIDAAKLKLATIA--AIGEIKTETTETFRFYVDYDDLMSLLK                | 275 |
| Db | 379 | IDKLKRDLMARDKNGIVLABETTYGEITLKLESQNRRELNMKMLLKALK              | 427 |

```

RESULT 30
US-10-270-085-8
; Sequence 8, Application US/10270085
; Patent No. 6627408
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Ronan
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/10/270,085
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/723,820
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-270-085-8

```

RESULT 31  
US-09-538-092-914  
; Sequence 914, Application US/09538092

```

; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 914
; LENGTH: 2349
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P12270
US-09-538-092-914

Query Match          6.8%; Score 99.5; DB 4; Length 2349;
Best Local Similarity 18.5%; Pred. No. 2.5;
Matches 60; Conservative 49; Mismatches 104; Indels 111; Gaps 11;

QY 27 YNKYL-DQVIPWKTTFDETIKELSRFKQEYSQ-----EASVL 61
DB 1354 YRKLLSEKVEHTKRIQDLTEEIGRLKAEIARSNASLTNNQNLQSLKEDLNKVRTEKETI 1413
QY 62 VGDIKVLLMDSQDKYEATQVYEWGVVTQLLSAYILLFDEYNEKKASQAQDILIRILD 121
DB 1414 QKDLDKAIIDIQEKVITIQ-----VKKIGERY---KTQTEELKAQODKVMETSQAQ 1462
QY 122 DG-----VNKLNBAQSKLLSGSSFNNASGKLIALDSQLTN-----DPS 160
DB 1463 SGDHQEHQSVQEMQELKETL-----NQAETKSKSLESQVENLQKTLSEKETEARNLQ 1515
QY 161 EKSSYQSQVDRTRKAYAGAAAGIYAGPGLIISYISAGVIEGKLIPELNDRLKAVQN 220
DB 1516 EQTVLQSELRLURQ-----DLQDRTTQEEQ 1541
QY 221 FFTSLSVTVKQANKDIDAAKLKLATEIAAI-----GEIKTETETTFRYVDYDDLMLS 272
DB 1542 LRQIITEKEKTKAIVAKSKIA-HLAGVKDQLTKENBELKQRNGALDQDKDELVDVRI 1600
QY 273 LLK----GAAKGMINTCNEYQQRH 292
DB 1601 ALKSQVEGRISRLRELREHQERH 1624

RESULT 32
US-09-134-000C-4522
; Sequence 4522, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4522
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; FEATURE:

```

NAME/KEY: MISC FEATURE  
 LOCATION: (8)-(8)  
 OTHER INFORMATION: Amino acid 8 is Xaa wherein Xaa = any amino acid.  
 US-09-134-000C-4522

Query Match  
 Best Local Similarity 6.5%; Score 98; DB 4; Length 618;  
 Matches 55; Conservative 51; Mismatches 104; Indels 54; Gaps 10;

91 TOLLSAYILLFD-EYNEKA---SAQKDL-IRILD----GVNKLNEAQSILGSSQS 140  
 72 TQMSNKLIVSDVLANLKALSDSNADISARIVDDKIGIRATNDLN--QONLICKND 129  
 141 FNN-----ASGKLALDSQ-----LTNDFSEK-----SSYFQSDVRIRKEAYAG 180  
 130 YRLNDFSTSKYQALNDKRVVNVVQPSQGTGTVGLVYKSNLENKYQEITNTASIF 189  
 181 AAGIVAGPFGILIIISYIAAGVIE-----GKLIPELNDRLKAVQNF 222  
 190 FTASIIIAAISIIVTLIARSITKPIGEMRQOIRIARGDVAGKVEVHGDELQLAET 249  
 223 TSLSVTVKQAKDIDAALKLATEIAAICEIKTETETTRFYVDYDDLMLSLKGAKKMI 282  
 250 NQLSERIEAQTMEARNLDSVLTHMTDGVIAIDRRGKVITINEMALSLLNVKNENVI 309  
 283 NT-----CNEYQORHGKTKLLLEVPD 302  
 310 GTSILLELDIEEDYTLRKLEBPD 333

RESULT 33  
 US-09-489-039A-8554  
 Sequence 8554, Application US/09489039A  
 Patent No. 6610836  
 GENERAL INFORMATION:  
 APPLICANT: Gary Breton et. al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 FILE REFERENCE: 2709.2004001  
 CURRENT APPLICATION NUMBER: US/09/489, 039A  
 PRIOR FILING DATE: 2000-01-27  
 PRIOR FILING DATE: 1999-01-29  
 NUMBER OF SEQ ID NOS: 14342  
 SEQ ID NO 8554  
 LENGTH: 1122  
 TYPE: PRT  
 ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-8554

Query Match  
 Best Local Similarity 6.5%; Score 98; DB 4; Length 1122;  
 Matches 68; Conservative 57; Mismatches 109; Indels 96; Gaps 14;

8 QTVFVWKS---AIFTADGALDFYKYLDOVIPW-KTFDETIKLSRPFKORYQASVLVG 63  
 57 ETVEVLQALNALNEERKSLERARQYQVDIDNFKLQSLRAQNLNLSSEPRQVPTGLTA 116  
 64 DIKVLMDSDQKYFEATQTVYEWCGVVTOLLSAVILLFDEYNEKKSAAOKDILIRILD-- 121  
 117 D-----ALNQELIQ-----VSSQLLE-----SSRQOQEQDRAREIADSL 151  
 122 -----DGVNKLNEAQSILGSSQSFNNA-----SGKILALDSQLTNDFSE 161  
 152 NQLPQQQTDAARQLNEVERRI--GTQTGNALAAQNALQASARLKAIVDEL--DLAQ 207  
 162 KSSVFQSDVRIRKEAYAGAGIAGVAGPFGILIIISYIAAGVIEGKLIPELNDRLKAVQNF 221  
 208 LSANNROELSRARSE-----LAQKQSEQLDAYLQALNRL 241  
 222 FTLSVTVKQAKDIDAALKL-----KIATRIAAIGEIKTE-TETTRFYVDYDDLMLSL 274  
 242 --QNSQOREAKALESTELLAENSENLPDITTAQFKVNRLESLQALNQQAQRMDLVAQ 299

275 KGA-----AKKMINTCNEYQORHGKTKLL 298  
 300 ROATNQTLOVQALNTLREQSOWLGSSNLL 329

RESULT 34  
 US-09-540-236-2753  
 Sequence 2753, Application US/09540236  
 Patent No. 6673910  
 GENERAL INFORMATION:  
 APPLICANT: Gary L. Breton et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARI  
 FILE REFERENCE: 2709.2005-001  
 CURRENT APPLICATION NUMBER: US/09/540,236  
 CURRENT FILING DATE: 2000-04-04  
 NUMBER OF SEQ ID NOS: 3840  
 SEQ ID NO 2753  
 LENGTH: 718  
 TYPE: PRT  
 ORGANISM: M. catarrhalis  
 US-09-540-236-2753

Query Match  
 Best Local Similarity 6.5%; Score 97.5; DB 4; Length 718;  
 Matches 66; Conservative 53; Mismatches 142; Indels 69; Gaps 11;

6 AEQTVFVWKSJETAADGALDFYKYLDOVIPWKTETIKLSRPFKOE-----YSQE 57  
 103 ATEKLEVAKEATQ-----DKVEKTSLSVEDIKDKAQSILQEDAADTVEALKQA 149  
 58 ASVLVGDIKY---LLMDSQDKYFEATQTVYEWCGVVTOLLSAVILLPEYVNC-----E 106  
 150 ASDKVETTKARAQSLKDDATQTFESAKQAVEGKVEAIKEQVLDQVDSLKDDTDQONTOD 209  
 107 KKAQAKDILIRILDGVNKLNEAQSILGSSQSF--NNASGKL-----LALD--SOLTND 158  
 210 QEKQTKDKAVQAATAAKRKVEDVDVHTVFESFKNTASGKIDKIKQAAVDEKTEEVKSQ 269  
 159 FSEKSSYFQSDVRIRKEAYAGAGIAGVAGPFGILIIISYIAAGVIEGKLIPELNDRLKAV 218  
 270 LSKADALKSSGEBELKQTAQTAANDAITAQ-AAVVGSGVAADSAQSAQSAKDKL---- 325  
 219 QNFTLSVTVKQAKDIDAALKLATEIAAICEIKTETETTRFYVDYDDLMLSLKGA 278  
 326 -----NQLFEOGKSALDEKQVLEKFGATEKINAVSENVDLATQVVIKEA 371  
 279 KKMINTC-----NEYQORHGKTKLL 297  
 372 QALQTNQAESLQAKAAGEEYDATHEDKGL 401

RESULT 35  
 US-08-785-428-2  
 Sequence 2, Application US/08785428  
 Patent No. 5750387  
 GENERAL INFORMATION:  
 APPLICANT: Hodgson, John  
 APPLICANT: Lawlor, Elizabeth  
 TITLE OF INVENTION: No. 5750387el tRNA Synthetase  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: SmithKline Beecham Corporation  
 STREET: 709 Swedeland Road  
 CITY: King of Prussia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19406-0939  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS





GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5178  
; LENGTH: 1010  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5178

Query Match 6.5%; Score 97.5; DB 3; Length 1010;  
Best Local Similarity 18.6%; Pred. No. 1.1; Mismatches 106; Indels 129; Gaps 14;  
Matches 66; Conservative 54

QY 38 KTFDETIKLSRF-----KQSYQASVVLGDIKVLMDSDQK-----YFEATQTVYEW 86  
DB 459 KLFDKQLDKSSYLSKLEKEKEQLNEISSERTINATLIDLNDKDFVNEIKSAMSGDT 518  
QY 87 CGV-----VTQLLSAYILLFDEY----- 104  
DB 519 CFIGCNEIHLGEHIDFESIAQKNNKIKLESKKVIRDEIKIETRIEELNHNRENELNF 578  
QY 105 --NEKK--ASAKDILIRLDGKVNKLENAQKSLGSSQSFNNAS-----GKLLALDSQLT 156  
DB 579 EKQEKDISELQKQ-----NHLNQLKDEQOSINKLVENFEKQKEIVNKIHQFDLDS 632  
QY 157 -----NDFSEKSSY-----FQSQVDRIKQAYAGAAAGIVAGPFGLIISYI 198  
DB 633 RKNQXKELEIQINDEPERSQSSYNDPETYSHAKQ-----VETVEY 676  
QY 199 AAGVIEGKLIPELNDRLKAVQNFSTLSVTVKQANKDIDAOKLKLATEIAAIGETETE 258  
DB 677 ENKTKDKL-NELNKKIEMNDQKHLTENLTQTSKEINNLEKKEKMQQLG----- 728  
QY 259 TTRFYVDYDLM-----LSLLKGAAGKMINTCNEYQQRH-----GKKTL 297  
DB 729 -----FESYDQVKSADLSAQKQEIETREINLYNKQSYEIEINRLKELVKGKLL 779

RESULT 38  
US-09-914-259-19  
; Sequence 19, Application US/09914259  
; Patent No. 6495336  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/09/914,259  
; CURRENT FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 975  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-914-259-19

Query Match 6.4%; Score 97; DB 4; Length 975;  
Best Local Similarity 21.1%; Pred. No. 1.1; Mismatches 118; Indels 62; Gaps 14;  
Matches 63; Conservative 56

QY 37 WKTDFETIKLSRPFQSYQASVVLGDIKVLMDSDQKFEATQTVYEWCGVVTQL--- 93

Db 376 WRA-GETVKAERQINMEDIMEAS-----TPNLEVEAAQTAATAAALAAQRTAL 422  
QY 94 --LSAYILLFDEYNEK-KASAKDILIRLDGKVNKLENAQKSLGSSQSFNNASGKLLA 150  
DB 423 ANMSASVAV-----NEQARLATECERLYOQLDDKBEINQ-----QSYAEQLKEQVME 471  
QY 151 LDSQLTNDPSEKSSYFQSQVDRIKQAYAG--AAGIVAGPFGLIISYIAAGVIEGKL- 207  
DB 472 QBELIANARREYET-LQSEMARIQENESAKEVEKVLQALELAVNYDQKSQEIDKNK 530  
QY 208 -IPELNDRLKAVQNFSTLSVTVKQANKDIDAOKLKLATEIAA-----IGEI----- 253  
DB 531 DIDALNEELQKQSVFNAASTELOQ-KDMSHQKRRITEMLTNLLRLDGEVGOAIPGE 589  
QY 254 -----KTETETTRFYVDYD--DLMLSLKGAAGKMINTCN--EYQQRHGKTKTLE 299  
DB 590 SSIDLKMSALAGTDASKVEEDFTMARLFISKMTAKNIAQRCNSMNETQOADSNNKISE 648

RESULT 39  
US-09-248-796A-19212  
; Sequence 19212, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 19212  
; LENGTH: 317  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-19212

Query Match 6.4%; Score 96.5; DB 4; Length 317;  
Best Local Similarity 19.1%; Pred. No. 0.22; Mismatches 53; Conservative 53; Indels 77; Gaps 10;  
Matches 53

QY 8 QTVFVWKSALETADGALDFYKYLQVLPWKTDFETIKLSRPFQSYQASVVLGDIKV 67  
DB 34 QKISAIKEKLE-----MDSN-----DESSRNKLQYISSYNRKLETKLRELKV 76  
QY 68 L---LMSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAQK-----DILIR 118  
DB 77 MQERLTKEHDDYRQEK-----ALLAKLYDIDFLYNETQEQVANLTKYAEITDLSILVK 129  
QY 119 ILDDGVNKLNEAQKSLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKQAY 178  
DB 130 ITEDLSNDLSMQTSYKNEKSQYQNG--ITIDSIQVNEINLK-----DLLKLAED-- 178  
QY 179 AGAAAGIVAGPFGLIISYIAAGVIEGKLIP-----ELNDRKAVQNFST 224  
DB 179 -----ITRNKQKQIDGKILKNDQNDDELTTKQMELEGQILIVODQLNK 223  
QY 225 LS-----VTVKQANKDIDAOKLKLATEIAAIGETETE 258  
DB 224 LNRKYQDILIVQADKKESAVKLEETKQEFNDILSEKE 261

RESULT 40  
US-09-248-796A-19055  
; Sequence 19055, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

;; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 107196.132  
;; CURRENT APPLICATION NUMBER: US/09/248,796A  
;; PRIOR FILING DATE: 1999-02-12  
;; PRIOR APPLICATION NUMBER: US 60/074,725  
;; PRIOR FILING DATE: 1998-02-13  
;; PRIOR APPLICATION NUMBER: US 60/096,409  
;; PRIOR FILING DATE: 1998-08-13  
;; NUMBER OF SEQ ID NOS: 28208  
;; SEQ ID NO 19055  
;; LENGTH: 858  
;; TYPE: PRT  
;; ORGANISM: Candida albicans  
;; FEATURE:  
;; NAME/KEY: UNSURE  
;; LOCATION: (851)  
;; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown

US-09-248-796A-19055

Query Match 6.4%; Score 96.5; DB 4; Length 858;  
Best Local Similarity 20.0%; Pred. No. 1;  
Matches 56; Conservative 46; Mismatches 107; Indels 71; Gaps 9;  
  
Qy 8 QTEVVKSAIETADGALDFYNYK--LDQVVPKTFDETIKELSRPKQEYSQEA5VLVCDI 65  
Db 489 QLVQEQINESIRIKNNFQIMSKFNQLRQSLSEKENDKNLDFSKNNHQOQOQOIQOLEQ 548  
  
Qy 66 KV-----LMDSDQKYFEATQVYEWCGVVTQLLSAYILLFDEYNEKKSAAKQDILI 117  
Db 549 KLIYEKCIITILQELDQYKPSDTTNTNNNNNNNNSSSYNNHNRSSLSNELNL 608  
  
Qy 118 -----RIIDDGVNKLNEAQKSLGSS--QSFNNASGKLALDLSQLTNDFSEKSS 164  
Db 609 VNDYLQQLTSYRINDELQVNDYK--LLNSSTLEKINN-----LTSKLQEKSI 656  
  
Qy 165 YFQSQVDRIRKEAYAGAAAGIVAGPFGLLISYSTAAGVIEGKLPELNDRLKAVQNFT- 223  
Db 657 ELRNQMG-----LNNKLQTELNLSDLKQRYNTE 685  
  
Qy 224 --SLSVTVVKQANKDIDAAKLKL--ATEIAAIGEIKTETET 259.  
Db 686 RIQSYTVDSLRKNEALQKLVNKLTLMTIDTRFAESNS 725

RESULT 41  
US-08-592-126-148  
; Sequence 148 Application US/08592126  
; Patent No. 5821091  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Dolganov  
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
; NUMBER OF SEQUENCES: Polypeptides  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/592,126  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0111

;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 324-0880  
;; TELEFAX: (415) 324-0960  
;; INFORMATION FOR SEQ ID NO: 148:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1312 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: NO  
;; ORIGINAL SOURCE:  
;; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54  
;; US-08-592-126-148

Query Match 6.4%; Score 96.5; DB 2; Length 1312;  
Best Local Similarity 18.7%; Pred. No. 2.1;  
Matches 74; Conservative 54; Mismatches 99; Indels 169; Gaps 16;  
  
Qy 10 VEVVKSIAETADGALDFYNYKLDQVVPKTFDETIK-----ELSRFKQEY 54  
Db 502 MEVISLQNEKAD--LDRTLRLKLDQEMEQLNHHHTTTTQOMELTKDKADKQEIIRKISRH 559  
  
Qy 55 SQEASVLVGDIKVLLMDSDQKYFEATQVYEWCGVVTQLLSAYILLFDEYNEKKSAAKQD 114  
Db 560 SDELTSLG-----YFPNKKQLEDWLLHLSK------EINQTR----- 591  
  
Qy 115 ILIRILDDGVNKLNEAQKSLGSSQSFNNASGKLALDLSQLTNDFSEK-----SSYFQS 168  
Db 592 -----DRLAKLN--KELASSEQNNHNNELKRREEQLSS-YEDKLPDVCQSQDFES 640  
  
Qy 169 QVDRIRKE-----AYAGAAAGIVAGP------ 190  
Db 641 DLRLKEIEKSKQRAMLGATA--VYSQFITQLTDENQSCCPVCQVRFQTEAELQEV 698  
  
Qy 191 -----GLI--ISYSTAAGVIEGKLPELNDRLKA 217  
Db 699 SDLOSURLAPDKLSTESLKKKEKRDDEMLGLVPMRQSIID--LKEKEIPELRNKLQN 756  
  
Qy 218 VQNFFTSLSVTVVKQANKDIDAAKLKLATEIAAIGEIKTETET-----RFYVDY 266  
Db 757 V-----NRDIQRLKNDIEQEETLLGTIMPEESAKVCLDVTIMERFQEL 802  
  
Qy 267 DDLMLSLKGA------KMINTCNEYQORHKK 295  
Db 803 KDVVERKIAQQAALQGLDLDRTVQVQVQKQEKQHK 838

RESULT 42  
US-08-687-080-51  
; Sequence 51, Application US/08687080  
; Patent No. 5965427  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Dolganov  
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 175  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/687,080  
; FILING DATE: 17-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/592,126

FILING DATE: 26-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0111.30  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1312 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NT.  
INDIVIDUAL ISOLATE: 389 TO 4324  
US-08-687-080-51

Query Match 6.4%; Score 96.5; DB 2; Length 1312;  
Best Local Similarity 18.7%; Pred. No. 2.1;  
Matches 74; Conservative 54; Mismatches 99; Indels 169; Gaps 16;  
QY 10 VEVVKSATADGALDFYNYKLDQVTPKTFDETIK-----ELSRFKQY 54  
Db 502 MEVISLQNEKAD--LDRTLKLDQEMQLNHHHTTTRTQMEMLTKDKADQDEQIRKISRH 559  
QY 55 SQASVLVGDIVKLLMDSQDKYFEATQTVVCGVVTQLLSAYILLDFEYNEKKASAKD 114  
Db 560 SDELTSLG-----YFPNKKQLEDLHLSKSK-----EINQTR----- 591  
QY 115 ILIRILDGYNKLENAQKSLGSSQFNNAAGKLLALDSQLTNDPSEK-----SSYFQS 168  
Db 592 -----DRLAKLN---KELASSEQNKNNHNNELKREQLSS--YEDKLFVCGSQDFES 640  
QY 169 QVDRIRKE-----AYAGAAGIVAGPF----- 190  
Db 641 DLRLKEEIEKSKORAMLAGATA--VYSOFITQLTDENQSCCPVQCFVTEAELOEVI 698  
QY 191 -----GLI-ISYSIAAGVIEGKLIPELNDRLKA 217  
Db 699 SLDQSKRLAPDKLKSTESLKKKRRDEMGLVPMRQSIID--LKEKEIPELNRKLQN 756  
QY 218 VONFFTSLSVTKQANKDIDAALKLATEIAAIGEIKTETETTT-----RFYVDY 266  
Db 757 V-----NRDIQRLKNDIEEQETLLGTIMPEESAKVCLTDTVIMERFQMEI 802  
QY 267 DDLMLSLKGAAK-----KMINTCNEYQQRHGKK 295  
Db 803 KQVERKIAQQAQAKLQGLDLDRTVQOVNQEKQKH 838

RESULT 43  
US-09-168-595-148  
Sequence 148, Application US/09168595  
Patent No. 655566  
GENERAL INFORMATION:  
APPLICANT: Gregory Dolganov  
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
TITLE OF INVENTION: Polypeptides  
NUMBER OF SEQUENCES: 151  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM: Floppy disk  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/168,595  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/592,126  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0111  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 148:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1312 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54  
US-09-168-595-148

Query Match 6.4%; Score 96.5; DB 4; Length 1312;  
Best Local Similarity 18.7%; Pred. No. 2.1;  
Matches 74; Conservative 54; Mismatches 99; Indels 169; Gaps 16;  
QY 10 VEVVKSATADGALDFYNYKLDQVTPKTFDETIK-----ELSRFKQY 54  
Db 502 MEVISLQNEKAD--LDRTLKLDQEMQLNHHHTTTRTQMEMLTKDKADQDEQIRKISRH 559  
QY 55 SQASVLVGDIVKLLMDSQDKYFEATQTVVCGVVTQLLSAYILLDFEYNEKKASAKD 114  
Db 560 SDELTSLG-----YFPNKKQLEDLHLSKSK-----EINQTR----- 591  
QY 115 ILIRILDGYNKLENAQKSLGSSQFNNAAGKLLALDSQLTNDPSEK-----SSYFQS 168  
Db 592 -----DRLAKLN---KELASSEQNKNNHNNELKREQLSS--YEDKLFVCGSQDFES 640  
QY 169 QVDRIRKE-----AYAGAAGIVAGPF----- 190  
Db 641 DLRLKEEIEKSKORAMLAGATA--VYSOFITQLTDENQSCCPVQCFVTEAELOEVI 698  
QY 191 -----GLI-ISYSIAAGVIEGKLIPELNDRLKA 217  
Db 699 SLDQSKRLAPDKLKSTESLKKKRRDEMGLVPMRQSIID--LKEKEIPELNRKLQN 756  
QY 218 VONFFTSLSVTKQANKDIDAALKLATEIAAIGEIKTETETTT-----RFYVDY 266  
Db 757 V-----NRDIQRLKNDIEEQETLLGTIMPEESAKVCLTDTVIMERFQMEI 802  
QY 267 DDLMLSLKGAAK-----KMINTCNEYQQRHGKK 295  
Db 803 KQVERKIAQQAQAKLQGLDLDRTVQOVNQEKQKH 838

RESULT 44  
US-09-248-796A-18018  
Sequence 18018, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstein et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725

```
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18018
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18018

Query Match      6.4%; Score 96; DB 4; Length 360;
Best Local Similarity 24.6%; Pred. No. 0.3;
Matches 46; Conservative 32; Mismatches 63; Indels 46; Gaps 8;

Qy 8 QTVEVVKSAIETADGAL----DFYNYKLDQVPIPKTFDETIKELSRFKQYSQBASVLVG 63
Db 54 QTIHEVKQAVASTDKATSEKDDSLKKHQDFLAKHKVHRHKLKQL-----EAE----- 100
Qy 64 DIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAS---AQKDILIRIL 120
Db 101 --EPLLVENKRYVMFPPIRYHE-----IWNFYKAAEASFMTAEEDLSKDL 144
Qy 121 DDGVNKLNEAKSLGSSQSFNNASGKLLALDSQLTNDPS-----EKSSY-FQSQVDR 172
Db 145 DDWNNKLNERYFTISRVLAFAASDGIWG--ENLIENFSTEVQLPEAKSFYGFQIMMEN 202
Qy 173 IRKEAYA 179
Db 203 IHSETYS 209

RESULT 45
US-09-543-681A-7696
; Sequence 7696, Application US/09543681A
; Patent No. 6505709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7696
; LENGTH: 1086
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7696

Query Match      6.4%; Score 96; DB 4; Length 1086;
Best Local Similarity 19.8%; Pred. No. 1.7;
Matches 48; Conservative 56; Mismatches 76; Indels 62; Gaps 12;

Qy 88 GVTQLLSAYILLFDEYNEKKASAKOIL-----IRILDDGVNKLNEAKSLGSSQSFN 142
Db 421 GIATGLLTAGAMAYEWQQAQESQKALEYAESLDVATESLNKLSDAERQA----- 472
Qy 143 NASGKLL-ALDSQLTNDPSKSSVFSQSDVIRKEAYAGAAGIVAGPFGLIISY----- 196
Db 473 -SIGKLAELGDAQI-----EKIEELKRQEDLKKSSDLGGANK-----GISFYTTGDYE 520
Qy 197 SIAGVIEGKLIPELNDRLKAVQNFFTSLSVTVVKQAKDIDAARKLKLATEIAAIGEIKTE 256
Db 521 SFSAEVVKAR-----KDELQVTADLETK-----NRELEVTRKKMALLLDA--EVQKN 565
Qy 257 TETTRFYVDY-----DDLMLSL-LKGAA-----KGMINTCNEYQQRHGKKTLLLE 300
Db 566 GWSNAYALYASRIYDASVKSEDLIGLKLHQSAPDNLQAQSRATQEQQFSQAQLIVI 625
Qy 301 PD 302
|
```

```
Db 626 SD 627

RESULT 46
US-09-538-092-1077
; Sequence 1077, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqformatter Version 0.9
; SEQ ID NO 1077
; LENGTH: 1960
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P35579
US-09-538-092-1077

Query Match      6.4%; Score 96; DB 4; Length 1960;
Best Local Similarity 19.1%; Pred. No. 4.4;
Matches 64; Conservative 48; Mismatches 133; Indels 90; Gaps 11;

Qy 7 EOTVEVVKSAIE---TADGALDFYNYKLDQVPIPKTFDETIKE-----LSRFEKQYS 55
Db 1166 EQEVNILLKKTLEEAKEAKTHEAQIQEMRQKHSQAV--EELAEQLEQTVRKVANLEKAKQTLE 1223
Qy 56 QEASVVLGVDIKVLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKQDI 115
Db 1224 NERGELEANEVKVLQGGKD-----SEHKRKKVEAQLE 1256
Qy 116 LIRILDDGVNKLNEAKSLGSSQSFNNASGKLLALD---SQTNDPSKSSVFSQSDVDR 172
Db 1257 LQVFNIEGERTVTELADKVTKLQVELDNVTGLLSQSDSKSLTKDFSALESQLODTQEL 1316
Qy 173 IRKEAYAGAAGIVAGPFGLIISYIAAGVIEGK-----LI 208
Db 1317 LOEENRQK-----LSLSTKLKQVEDEKNSPREOLEEEEAHKNLEKQIATLH 1363
Qy 209 PELNDRLLKAVQNFFTSLSV---TVKQAKDIDAARKLKLATEIAAIGEIKTETTRFYVD 265
Db 1364 AQVADMKKMKEDSVGCLTAEAEVVKRKLQKDLGLSQRHEEKVAAYD--KLEKTKRLQOE 1421
Qy 266 YDDLMLSLKGAACKMINTCN-EYQQRHGKKTLLLE 299
Db 1422 LDDELLVDL---DHQRSACNLEKKKKQKFDQLLAE 1452

RESULT 47
US-09-538-092-936
; Sequence 936, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
```

[illegible]

RESULT 49  
 ; US-08-434-000A-4  
 ; Sequence 4, Application US/08434000A  
 ; Patent No. 6046037  
 ;  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ANDREW C. HIATT, JULIAN  
 ; APPLICANT: K.-C. MA, THOMAS LEHNER  
 ; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
 ; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: 633 West Fifth Street  
 ; STREET: Suite 4700  
 ; CITY: Los Angeles  
 ; STATE: California  
 ; COUNTRY: U.S.A.  
 ; ZIP: 90071  
 ;  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; MEDIUM TYPE: storage  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0  
 ; SOFTWARE: Word Perfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/434,000A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; PRIOR APPLICATION DATA: including application  
 ; PRIOR APPLICATION DATA: described below:  
 ; APPLICATION NUMBER: 08/367,395  
 ; FILING DATE: 12/30/94  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Guise, Jeffrey W.  
 ; REGISTRATION NUMBER: 34,613  
 ; REFERENCE/DOCKET NUMBER: 212/127  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 552-8400  
 ; TELEFAX: (619) 552-0159  
 ; TELEX: 67-3510  
 ; TELEX: SEQUENCE LISTING  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 746 amino acids

```

; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 936
; LENGTH: 2871
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P15924
US-09-538-092-936

Query Match
Best Local Similarity 17.6%; Pred. No. 8;
Matches 59; Conservative 56; Mismatches 92; Indels 128; Gaps 12;

QY 28 NKYLQVWPVKTFDETIKELSRFQESQBSASVLVG-----DIKVLMDSQDKYFEAT 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1208 NKYEINITKT---TIKEISMQKEDDSKNLRNQDLRSRENRLDKDEIVRLNDSILOAT 1264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 81 Q-----TVYEWCGVVTQLLSAYILLDFEYNEKASQAQ---KDILIRILDDGV--N 125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1265 EQRRRAEENALQOKAC-----SEIMQKQHLEILKQVMOQRSEDNARHKQ 1311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 126 KLINEAKSLGSSQSF-----NNASGKLLALDS----- 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1312 SLEBAKTIQDKKEIERLKAEFQEAERWEYENELSKVRNNVDEIISLKNQPETFIN 1371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 154 -----QLTNDPFSKSYFSQVDRIKRAYAGAAAGIVAGPGLIISYSIAAGVIEGK 206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1372 ITKTIHQLTWQKEEDTSGYRAQIDNLTRENRS----- 1404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 207 LPELNDRLKAVQNFETSLSVTVKQAKNDIDAACL-----KLATEIAAIGBIKTE 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1405 -----LSEETIKRLKTLTOTTENLRVVEDIOQQKATSEVSQRKQQLLEVELRQVTOMRTE 1460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 257 TETTRPYVDYDDLMLSLKGAAGKAGMINTCNEYQQR 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1461 -ESVRVYKQSLDD-----AAKTIDKNKEIER 1485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 48
US-08-875-435B-3
; Sequence 3, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
; APPLICANT: Hasegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji
; APPLICANT: Matsuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro
; APPLICANT: Iehiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; TITLE OF INVENTION: RECOMBINANT DNA
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875,435B
; CURRENT FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-875-435B-3

Query Match
6.3%; Score 95.5; DB 4; Length 1972;

```

```
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Human Polyimmunoglobulin Receptor
US-08-434-000A-4
Query Match 6.3%; Score 94.5; DB 3; Length 746;
Best Local Similarity 20.7%; Pred. No. 1.3;
Matches 41; Conservative 47; Mismatches 69; Indels 41; Gaps 11;
QY 86 WCGV-----VTQLLSAYILLFDEYNEKKASAKDILI-----RILDDGVNKLNEAQ 131
Db 525 WCGVKQGHFYGETAAVYVAV-----EERKAAGSRDVS LAKADAAPEKVLDSG---FREIE 577
QY 132 KSLIGSSQSFNNASCKLLALDSQLTNDPSEKSSYFQSQVDRIKEAYAGAAAGIVAG--P 189
Db 578 NKAIQDPRLP--ABEKAVA-----DTRDAQDSRASVDSSGSEEGGSSRALVSTLVP 628
QY 190 FGLIISY-SIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQ-----ANKDIDAALK 242
Db 629 LGLVLAVGAVAGVARARHRKNV-DRV-SIRSYRTDISMSDFENSREFGANDNMGCASSIT 686
QY 243 LATEIAAIGEIKTETETT 260
Db 687 QETSLGGKEEFVATTEST 704

RESULT 50
US-09-312-157-4
; Sequence 4, Application US/09312157
; Patent No. 6303341
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/312,157
; FILING DATE: 14-May-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/434,000
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-351
; SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 746 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; DESCRIPTION: Human Polyimmunoglobulin Receptor
; SEQ ID NO: 4
;
; Query Match 6.3%; Score 94.5; DB 3; Length 746;
; Best Local Similarity 20.7%; Pred. No. 1.3;
; Matches 41; Conservative 47; Mismatches 69; Indels 41; Gaps 11;
;
; QY 86 WCGV-----VTQLLSAYILLFDEYNEKKASAKDILI-----RILDDGVNKLNEAQ 131
; Db 525 WCGVKQGHFYGETAAVYVAV-----EERKAAGSRDVS LAKADAAPEKVLDSG---FREIE 577
; QY 132 KSLIGSSQSFNNASCKLLALDSQLTNDPSEKSSYFQSQVDRIKEAYAGAAAGIVAG--P 189
; Db 578 NKAIQDPRLP--ABEKAVA-----DTRDAQDSRASVDSSGSEEGGSSRALVSTLVP 628
; QY 190 FGLIISY-SIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQ-----ANKDIDAALK 242
; Db 629 LGLVLAVGAVAGVARARHRKNV-DRV-SIRSYRTDISMSDFENSREFGANDNMGCASSIT 686
; QY 243 LATEIAAIGEIKTETETT 260
; Db 687 QETSLGGKEEFVATTEST 704
;
; Search completed: January 28, 2005, 19:17:33
; Job time : 43 secs
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:03:34 ; Search time 157 Seconds  
(without alignments)  
692.325 Million cell updates/sec

Title: US-09-993-292B-24

Perfect score: 1510

Sequence: 1 MTGIFAEQTEVVKSAIETA.....TCNEYQORHGKTKLLEVPDI 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database :

A\_Geneseq\_23Sep04:\*

1: geneseq1980s:\*

2: geneseq1990s:\*

3: geneseq2000s:\*

4: geneseq2001s:\*

5: geneseq2002s:\*

6: geneseq2003as:\*

7: geneseq2003bs:\*

8: geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 1466  | 97.1        | 305    | 6     | ABG73265 Salmonell |
| 2          | 1466  | 97.1        | 749    | 6     | ABG73266 Salmonell |
| 3          | 1370  | 90.7        | 309    | 2     | AAR67645 Haemolysi |
| 4          | 1065  | 70.5        | 1035   | 4     | ABG25221 Novel hum |
| 5          | 126.5 | 8.4         | 582    | 6     | ABM68103 Phototrab |
| 6          | 116   | 7.7         | 941    | 6     | ABU18419 Protein e |
| 7          | 113   | 7.5         | 1015   | 6     | ABP76825 N. gonorr |
| 8          | 113   | 7.5         | 1015   | 6     | ABP76817 N. gonorr |
| 9          | 111.5 | 7.4         | 1036   | 7     | ADF07451 Bacterial |
| 10         | 110.5 | 7.3         | 2681   | 6     | ABJ19025 Pathogen  |
| 11         | 110   | 7.3         | 1189   | 6     | ABU42396 Protein e |
| 12         | 109   | 7.2         | 515    | 7     | ADC95690 E. faeciu |
| 13         | 108.5 | 7.2         | 1196   | 7     | ADC94317 E. faeciu |
| 14         | 107.5 | 7.1         | 742    | 6     | ABU25599 Protein e |
| 15         | 106.5 | 7.1         | 3692   | 6     | ABU43311 Protein e |
| 16         | 106.5 | 7.1         | 3696   | 5     | ABP40235 Staphyloc |
| 17         | 106   | 7.0         | 1231   | 7     | ADC95523 E. faeciu |
| 18         | 106   | 7.0         | 3187   | 7     | ADE56031 Rat Prote |
| 19         | 106   | 7.0         | 3187   | 7     | ADE56035 Rat Prote |
| 20         | 105.5 | 7.0         | 588    | 4     | ABB71280 Drosophil |
| 21         | 104.5 | 6.9         | 1292   | 5     | ABP77986 Amino aci |
| 22         | 104   | 6.9         | 685    | 4     | ABM64018 Drosophil |
| 23         | 104   | 6.9         | 722    | 4     | AAU35203 Enterococ |
| 24         | 104   | 6.9         | 815    | 6     | ADA32997 Acinetoba |
| 25         | 104   | 6.9         | 1211   | 5     | ABP39975 Staphyloc |

|    |       |     |      |   |                     |
|----|-------|-----|------|---|---------------------|
| 26 | 104   | 6.9 | 2334 | 5 | ABG31849 Human kin  |
| 27 | 103.5 | 6.9 | 541  | 5 | ABP29330 Streptoco  |
| 28 | 103.5 | 6.9 | 660  | 4 | AAU38921 C. tracho  |
| 29 | 103.5 | 6.9 | 660  | 7 | ADD42726 Chlamydia  |
| 30 | 103.5 | 6.9 | 1020 | 6 | ABU30092 Protein e  |
| 31 | 103.5 | 6.9 | 1163 | 6 | ABU23940 Protein e  |
| 32 | 103.5 | 6.9 | 1976 | 7 | ADE63514 Rat Prote  |
| 33 | 103.5 | 6.9 | 1976 | 7 | ADE63518 Rat Prote  |
| 34 | 103.5 | 6.9 | 5373 | 4 | AAU14603 Novel bon  |
| 35 | 103.5 | 6.9 | 5373 | 7 | ADJ68935 Human hea  |
| 36 | 103   | 6.8 | 284  | 7 | ADH85709 Enterococ  |
| 37 | 103   | 6.8 | 885  | 4 | AGH82283 S. epider  |
| 38 | 103   | 6.8 | 961  | 7 | ADB67135 General v  |
| 39 | 103   | 6.8 | 961  | 8 | ADJ69406 Nanostruc  |
| 40 | 103   | 6.8 | 1583 | 7 | ADJ69829 Human hea  |
| 41 | 103   | 6.8 | 2056 | 4 | ABB59344 Drosophil  |
| 42 | 102.5 | 6.8 | 472  | 6 | ABU27359 Protein e  |
| 43 | 102.5 | 6.8 | 746  | 3 | AGG46982 Arabidops  |
| 44 | 102.5 | 6.8 | 788  | 3 | AGG46981 Arabidops  |
| 45 | 102.5 | 6.8 | 5447 | 4 | AAU14697 Novel bon  |
| 46 | 102   | 6.8 | 590  | 4 | AAE10039 N. mening  |
| 47 | 102   | 6.8 | 590  | 4 | AAU27604 Neisseria  |
| 48 | 102   | 6.8 | 612  | 4 | AAE10042 N. mening  |
| 49 | 102   | 6.8 | 612  | 4 | AAU27607 Neisseria  |
| 50 | 102   | 6.8 | 739  | 4 | AAAB96493 Putative  |
| 51 | 102   | 6.8 | 1948 | 5 | ABP73774 Candida a  |
| 52 | 101.5 | 6.7 | 476  | 2 | AAAR43563 Hyaluroa  |
| 53 | 101.5 | 6.7 | 883  | 6 | ADB10082 Alloiooc   |
| 54 | 101.5 | 6.7 | 890  | 6 | ADB10080 Alloiooc   |
| 55 | 101.5 | 6.7 | 1857 | 5 | AAU84350 Protein M  |
| 56 | 101.5 | 6.7 | 1938 | 6 | ABR92127 Human cer  |
| 57 | 101.5 | 6.7 | 1945 | 6 | ABU10398 Human em   |
| 58 | 101.5 | 6.7 | 1972 | 6 | ABR92126 Human cer  |
| 59 | 101.5 | 6.7 | 1972 | 7 | ADE58724 Human Pro  |
| 60 | 101.5 | 6.7 | 1972 | 7 | ADE58721 Human Pro  |
| 61 | 101.5 | 6.7 | 1972 | 7 | ADE63520 Human Pro  |
| 62 | 101.5 | 6.7 | 1972 | 7 | ADE63516 Human Pro  |
| 63 | 101.5 | 6.7 | 1972 | 8 | ADN03745 Antipsoi   |
| 64 | 101.5 | 6.7 | 1972 | 8 | ADN03745 Antipsoi   |
| 65 | 101.5 | 6.7 | 1979 | 6 | ABU10399 Human em   |
| 66 | 101   | 6.7 | 1129 | 7 | ADF07734 Bacterial  |
| 67 | 100.5 | 6.7 | 425  | 8 | ADG98222 Human apo  |
| 68 | 100.5 | 6.7 | 472  | 2 | AAU37727 Protein i  |
| 69 | 100.5 | 6.7 | 529  | 6 | ABU20414 Protein e  |
| 70 | 100.5 | 6.7 | 553  | 3 | AAU97403 Human apo  |
| 71 | 100.5 | 6.7 | 566  | 6 | ABR53540 Protein s  |
| 72 | 100.5 | 6.7 | 566  | 7 | ADK62550 Disease t  |
| 73 | 100.5 | 6.7 | 609  | 7 | ADJ70327 Human hea  |
| 74 | 100.5 | 6.7 | 609  | 7 | ADJ68760 Human hea  |
| 75 | 100.5 | 6.7 | 613  | 3 | AAAB15566 Fragment  |
| 76 | 100.5 | 6.7 | 613  | 3 | AAAB15558 Apoptosis |
| 77 | 100.5 | 6.7 | 613  | 3 | AAU92512 Human OXR  |
| 78 | 100.5 | 6.7 | 613  | 5 | ABB89626 Human pol  |
| 79 | 100.5 | 6.7 | 613  | 5 | ADD25643 Binding d  |
| 80 | 100.5 | 6.7 | 613  | 7 | ADE48157 Human AIF  |
| 81 | 100.5 | 6.7 | 613  | 7 | ADJ68765 Human hea  |
| 82 | 100.5 | 6.7 | 623  | 3 | AAU97401 Human apo  |
| 83 | 100.5 | 6.7 | 627  | 3 | AAU97399 Human apo  |
| 84 | 100.5 | 6.7 | 803  | 4 | AAU34101 Staphyloc  |
| 85 | 100.5 | 6.7 | 805  | 4 | AAU36693 Staphyloc  |
| 86 | 100.5 | 6.7 | 861  | 8 | AAU48067 Polyptei   |
| 87 | 100.5 | 6.7 | 1009 | 6 | ABU15864 Protein e  |
| 88 | 100.5 | 6.7 | 1377 | 4 | ABB65439 Drosophil  |
| 89 | 100.5 | 6.7 | 7201 | 4 | ABB71136 Drosophil  |
| 90 | 100   | 6.6 | 490  | 8 | ADM72386 A. thalia  |
| 91 | 100   | 6.6 | 523  | 5 | ABB78800 Human NOV  |
| 92 | 100   | 6.6 | 927  | 5 | ABB49720 Listeria   |
| 93 | 100   | 6.6 | 990  | 7 | ADM05211 Human pro  |
| 94 | 100   | 6.6 | 1174 | 6 | ADA54343 Human pro  |
| 95 | 100   | 6.6 | 1886 | 2 | AAW54241 Rattus no  |
| 96 | 100   | 6.6 | 1939 | 5 | ABB77096 Human alp  |
| 97 | 100   | 6.6 | 2481 | 5 | ABG92080 Human rec  |
| 98 | 99.5  | 6.6 | 610  | 5 | ABG80347 Clostridi  |

99 99.5 6.6 1048 4 ABB59245 AbB59245 Drosophil  
100 99.5 6.6 1066 4 AAG67418 Aag67418 Amino aci

## ALIGNMENTS

RESULT 1  
ABG73265  
ID ABG73265 standard; protein; 305 AA.  
XX AC ABG73265;  
XX  
DT 17-APR-2003 (first entry)  
XX  
DE Salmonella typhi cytolsin A (ClyA) protein.  
XX  
KW Protein expression vector; expression cassette; export protein;  
KW transformed host cell; bacterial cell; immune response; HlyE family;  
KW cytolsin A; ClyA; recombinant protein; bioreactor; bacterial growth;  
KW live vector immunogenic composition; animal feed; animal vaccination;  
KW food industry; nutritional supplement; biomediation; waste disposal;  
KW waste treatment; host bacterium.  
XX  
OS Salmonella typhi.  
XX  
FN US2002146430-A1.  
XX  
PD 10-OCT-2002.  
XX  
PF 23-NOV-2001; 2001US-00993292.  
XX  
PR 22-NOV-2000; 2000US-0252516P.  
XX  
PA (GALE/) GALEN J E.  
XX  
PI Galen JE;  
XX  
DR WPI; 2003-228013/22.  
DR N-FSDB; ABX15174.  
XX  
PT Novel system for expressing protein of interest, has expression vector  
PT with cassette having export protein coding sequence fused to protein  
PT coding sequence, host cell transformed with vector, and culturing  
PT environment.  
XX  
PS Claim 5; Page 18; 33pp; English.  
XX  
CC The present invention relates to a system for expressing a protein of  
CC interest. The system comprises an expression vector having an expression  
CC cassette, where the expression cassette comprises an export protein  
CC coding sequence genetically fused to a DNA sequence encoding the protein  
CC of interest, a host cell transformed with the expression vector, and a  
CC culturing environment for the transformed host cell, where the expression  
CC cassette expresses the export protein-protein of interest fusion protein  
CC which is exported out of the host cell. The system is useful for  
CC expressing a gene in a bacterial cell, by providing an expression vector  
CC to a population of untransformed bacterial cells, and expressing the  
CC expression cassette, such that the export protein-protein of interest  
CC fusion protein is produced and exported into the culture medium. The  
CC protein of interest is preferably an antigen and this method is useful  
CC for eliciting an immune response from a host against the fusion protein.  
CC The export protein may be a member of the HlyE family such as cytolsin A  
CC (ClyA). The system is useful for the production of recombinant proteins  
CC inside a bacterial host cell, in a bioreactor, and various devices that  
CC facilitate the growth of bacteria. The system is also useful for  
CC providing an animal antigenic material against which an immune response  
CC may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can  
CC be used as a live vector immunogenic composition capable of facilitating  
CC the generation of an immune response in an animal. The system is also  
CC useful for preparing immunogenic compositions for vaccinating animals,  
CC and for use in the food, nutritional supplement, animal feed,  
CC biomediation, waste disposal, and waste treatment industries. The system

is very advantageous for protein expression, as proteins that are toxic  
to host bacterium can also be expressed. A population of recombinant host  
cells can be transformed with the expression vector, and then the  
population of bacterial host cells can be maintained in culture and used  
to produce proteins without having to harvest and lyse the host cells.  
The present represents Salmonella typhi ClyA protein

Sequence 305 AA;

Query Match 97.1%; Score 1466; DB 6; Length 305;  
Best Local Similarity 97.0%; Pred. No. 7.1e-114;  
Matches 294; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
QY 1 MTGIFAEQTVVEVKSAIETADGALDFYKYLQDVI PKMTFDETIKELSRPKQEYSQASV 60  
DB 1 MTSIFAEQTVVEVKSAIETADGALDFYKYLQDVI PKMTFDETIKELSRPKQEYSQASV 60  
QY 61 LVGDIKVLMDSDQDKYFEATQTVYVCGVVTQLLSAYILLFDEYNEKKASAKDILIRIL 120  
DB 61 LVGDIKVLMDSDQDKYFEATQTVYVCGVVTQLLSAYILLFDEYNEKKASAKDILIRIL 120  
QY 121 DDGVNKLNEAKSLGSSQSFNNASGKLLALDSQLTNDSEKSSYFQSDVDRIRKEAYAG 180  
DB 121 DDGVNKLNEAKSLGSSQSFNNASGKLLALDSQLTNDSEKSSYFQSDVDRIRKEAYAG 180  
QY 181 AAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNDRLKAVQNFPTSLSVTVKQANKDIDAAK 240  
DB 181 AAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNDRLKAVQNFPTSLSVTVKQANKDIDAAK 240  
QY 241 LKLAETIAAIGEIKTETETTRFYVDYDDLMLSLKGAAKMINTCNEYQQRHGKKTLLFV 300  
DB 241 LKLAETIAAIGEIKTETETTRFYVDYDDLMLSLKGAAKMINTCNEYQQRHGKKTLLFV 300  
QY 301 PDI 303  
DB 301 PDV 303

## RESULT 2

ABG73266  
ID ABG73266 standard; protein; 749 AA.

XX AC ABG73266;  
XX  
DT 17-APR-2003 (first entry)  
XX  
DE Salmonella typhi ClyA-Bacillus subtilis sacB fusion protein.

Protein expression vector; expression cassette; export protein;  
transformed host cell; bacterial cell; immune response; HlyE family;  
cytolsin A; ClyA; recombinant protein; bioreactor; bacterial growth;  
live vector immunogenic composition; animal feed; animal vaccination;  
food industry; nutritional supplement; biomediation; waste disposal;  
waste treatment; host bacterium; sacB; mutant; mutein;  
lethal levansucrase.

OS Salmonella typhi.  
OS Bacillus subtilis.  
OS Synthetic.  
OS Chimeric.  
PN US2002146430-A1.  
PD 10-OCT-2002.  
XX 23-NOV-2001; 2001US-00993292.  
XX 22-NOV-2000; 2000US-0252516P.  
XX (GALE/) GALEN J E.  
XX Galen JE;  
XX

DR WPI; 2003-228013/22.  
DR N-PSDB; ABX15191.

PT Novel system for expressing protein of interest, has expression vector  
PT with cassette having export protein coding sequence fused to protein  
PT coding sequence, host cell transformed with vector, and culturing  
PT environment.

XX Example 3; Page 25-28; 33pp; English.

XX The present invention relates to a system for expressing a protein of  
XX interest. The system comprises an expression vector having an expression  
XX cassette, where the expression cassette comprises an export protein  
XX coding sequence genetically fused to a DNA sequence encoding the protein  
XX of interest, a host cell transformed with the expression vector, and a  
XX culturing environment for the transformed host cell, where the expression  
XX cassette expresses the export protein-protein of interest fusion protein  
XX which is exported out of the host cell. The system is useful for  
XX expressing a gene in a bacterial cell, by providing an expression vector  
XX to a population of untransformed bacterial cells, and expressing the  
XX expression cassette, such that the export protein-protein of interest  
XX fusion protein is produced and exported into the culture medium. The  
XX protein of interest is preferably an antigen and this method is useful  
XX for eliciting an immune response from a host against the fusion protein.  
XX The export protein may be a member of the HlyE family such as cytolysin A  
XX (ClyA). The system is useful for the production of recombinant proteins  
XX inside a bacterial host cell, in a bioreactor, and various devices that  
XX facilitate the growth of bacteria. The system is also useful for  
XX providing an animal antigenic material against which an immune response  
XX may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can  
XX be used as a live vector immunogenic composition capable of facilitating  
XX the generation of an immune response in an animal. The system is also  
XX useful for preparing immunogenic compositions for vaccinating animals,  
XX and for use in the food, nutritional supplement, animal feed,  
XX biomediation, waste disposal, and waste treatment industries. The system  
XX is very advantageous for protein expression, as proteins that are toxic  
XX to host bacterium can also be expressed. A population of recombinant host  
XX cells can be transformed with the expression vector, and then the  
XX population of bacterial host cells can be maintained in culture and used  
XX to produce proteins without having to harvest and lyse the host cells.  
XX The present sequence represents Salmonella typhi ClyA-Bacillus subtilis  
XX lethal levanucrase (encoded by sacB gene) fusion protein. The fusion  
XX protein is used to investigate the versatility of ClyA as a fusion  
XX partner to export heterologous antigens out of S. typhi

XX Sequence 749 AA;

Query Match 97.1%; Score 1466; DB 6; Length 749;  
Best Local Similarity 97.0%; Pred. No. 2.3e-113;  
Matches 294; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 MTGIFAEQTVVVKSAIETADGALDFYKNYLDQVTPWKTFTDTIKELSRFKQEQYSQASV 60  
DB 1 MTSIFAEQTVVVKSAIETADGALDFYKNYLDQVTPWKTFTDTIKELSRFKQEQYSQASV 60  
QY 61 LVGDIKVLMDSDQKYFEATQTVYEWCGVWTQLLSAYILLFDEYNEKKASAKQDILIRIL 120  
DB 61 LVGDIKVLMDSDQKYFEATQTVYEWCGVWTQLLSAYILLFDEYNEKKASAKQDILIRIL 120  
QY 121 DDGVNKLNEAOKSLILGSSQSFNNASGKLLALDSQLTNDPFSKSSYFQSQVDRIKEAYAG 180  
DB 121 DDGVNKLNEAOKSLILGSSQSFNNASGKLLALDSQLTNDPFSKSSYFQSQVDRIKEAYAG 180  
QY 181 AAAGIVAGPFGILLISYSIAAGVIEGKLIPELNDRLKAVQNFSTLSVTVKQANKDIDA 240  
DB 181 AAAGIVAGPFGILLISYSIAAGVIEGKLIPELNNRLKTVQNFSTLSVTVKQANKDIDA 240  
QY 241 LKLAIEAAGISIKTETETTRFYVDYDDLMLSLKGAAKMINTCNEYQQRHGKKTILEV 300  
DB 241 LKLAIEAAGISIKTETETTRFYVDYDDLMLSLKGAAKMINTCNEYQQRHGKKTILEV 300  
QY 301 PDI 303  
|||

DB 301 PDV 303

RESULT 3

AAR67645

ID AAR67645 standard; protein; 309 AA.

XX AAR67645;

XX 25-MAR-2003 (revised)

XX 06-JUL-1995 (first entry)

XX Haemolysin regulator.

XX Haemolysin regulator; tuberculosis; vaccine; therapy.

XX Escherichia coli.

XX WO9428137-A1.

XX 08-DEC-1994.

XX 24-MAY-1994; 94WO-US005869.

XX 24-MAY-1993; 93US-00066830.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX King CH, Shinnick TM, Sathish M;

XX WPI; 1995-022809/03.

XX P-PSDB; AAQ75857.

XX New nucleic acid encoding haemolytic regulator of E coli - and derived  
XX vectors, proteins etc used to transform Mycobacterium cells to produce  
XX antituberculosis vaccines providing improved immune response.

XX Claim 9; Page 39-40; 46pp; English.

XX The sequence represents the E. coli haemolysis regulator which may be  
XX used as an immunogen in a vaccine composition against tuberculosis  
XX (Mycobacterium tuberculosis). (Updated on 25-MAR-2003 to correct PN  
XX field.) (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-  
XX 2003 to correct PA field.)

XX Sequence 309 AA;

Query Match 90.7%; Score 1370; DB 2; Length 309;  
Best Local Similarity 90.0%; Pred. No. 7.3e-106;  
Matches 271; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 1 MTGIFAEQTVVVKSAIETADGALDFYKNYLDQVTPWKTFTDTIKELSRFKQEQYSQASV 60  
DB 1 MTEIVADKTVVVKSAIETADGALDFYKNYLDQVTPWKTFTDTIKELSRFKQEQYSQASV 60  
QY 61 LVGDIKVLMDSDQKYFEATQTVYEWCGVWTQLLSAYILLFDEYNEKKASAKQDILIRIL 120  
DB 61 LVGDIKVLMDSDQKYFEATQTVYEWCGVWTQLLSAYILLFDEYNEKKASAKQDILIRIL 120  
QY 121 DDGVNKLNEAOKSLILGSSQSFNNASGKLLALDSQLTNDPFSKSSYFQSQVDRIKEAYAG 180  
DB 121 DDGVNKLNEAOKSLILGSSQSFNNASGKLLALDSQLTNDPFSKSSYFQSQVDRIKEAYAG 180  
QY 181 AAAGIVAGPFGILLISYSIAAGVIEGKLIPELNDRLKAVQNFSTLSVTVKQANKDIDA 240  
DB 181 AAAGIVAGPFGILLISYSIAAGVIEGKLIPELNNRLKTVQNFSTLSVTVKQANKDIDA 240  
QY 241 LKLAIEAAGISIKTETETTRFYVDYDDLMLSLKGAAKMINTCNEYQQRHGKKTILEV 300  
DB 241 LKLAIEAAGISIKTETETTRFYVDYDDLMLSLKGAAKMINTCNEYQQRHGKKTILEV 300  
QY 301 P 301  
|||

Db 301 P 301

RESULT 4  
 ABG25221  
 ID ABG25221 standard; protein; 1035 AA.  
 XX  
 AC ABG25221;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 XX Novel human diagnostic protein #25212.  
 DE  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS89408.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 55580; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological actions in  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1035 AA;  
 Query Match 70.5%; Score 1065; DB 4; Length 1035;  
 Best Local Similarity 91.3%; Pred. No. 9.8e-80;  
 Matches 210; Conservative 12; Mismatches 8; Indels 0; Gaps 0;  
 QY 70 MDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKILIRLDDGVNKLNE 129  
 DB 1 MDSQDKYFEATQTVYEWCGVATQLLAAYILLFDEYNEKKAQAQKILIRLDDGVNKLNE 60

QY 130 AQKSLIGSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSQVDRIKKEAYAGAAGIVAGP 189  
 DB 61 AQKSLVSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSQVDKIRKEAYAGAAGVAGP 120  
 QY 190 FGLIISYSTAAGVIECKLIPELNDRKAVONFTSLSVTKOANKDIDAOKLKLATEIAA 249  
 DB 121 FGLIISYSTAAGVIECKLIPELNDRKAVONFTSLSVTKOANKDIDAOKLKLATEIAA 180  
 QY 250 IGEIKTETETTRFYVDYDDMLSLKGAARKMINTCNEYQQRHGKKTLE 299  
 DB 181 IGEIKTETETTRFYVDYDDMLSLKGAARKMINTCNEYQQRHGKKTLE 230

RESULT 5  
 ABM68103  
 ID ABM68103 standard; protein; 582 AA.  
 XX  
 AC ABM68103;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Photorhabdus luminescens protein sequence #1200.  
 XX  
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN WO200294867-A2.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 07-FEB-2002; 2002WO-IB003040.  
 XX  
 PR 07-FEB-2001; 2001FR-00001659.  
 XX  
 PA (INSP) INST PASTEUR.  
 PA (CNRS) CNRS CENT NAT RECH SCI.  
 XX  
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 XX  
 DR WPI; 2003-148459/14.  
 XX  
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX  
 PS Claim 2; SEQ ID NO 1200; 1205pp; French.  
 XX  
 CC The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins

Query Match 8.4%; Score 126.5; DB 6; Length 582;  
Best Local Similarity 19.0%; Pred. No. 0.089;  
Matches 80; Conservative 55; Mismatches 133; Indels 155; Gaps 14;

1 MTGIFAEQTVGVK-----SAITADGALDFYKYLDOV-----IPWKT----- 40  
107 ITGDFSGKFSGLVNFPMKLPNRGDNLETPPPAYFLFYDQIKSWSSPMNSPENLQGY 166  
41 -----DETIKLSRFKQBYSOEASVVLGDKVL----- 68  
167 SNWKGPLIKYFTGYLKPEHFDIENIYYSIKESAHKIEKFOSAVEIVDNTVDSPT 226  
69 -----LMSDQKQYFEATQTVYEWGVVTVQLLSAYILLFDEYNKKA 109  
227 LDNNDFIKQNEIKNELYDLIDYQRLYDAQATI---TSNIYDLESQYALAITANELEA 283  
110 -----SAQDKIL-----TRILDDGVNKLNEAKSLGSSQSFNNSG 146  
284 DYKFAVESIPTDYLCPLCGTLHDNSLPNALLSEKDSLLNEA-----NSIAS 332  
147 KLLALDSQLTNDSEKSSVFOSQVDIRKEAVAGAAGIVAGPGLIISYIAAGVTEGK 206  
333 KIVELKSSL-NSLNEDAQFTNEIERNNK-----YITDDDAVKKG 372  
207 LIPELNDRL-----KAVQNFFTSLSVTVKQANKDI-----DAAKL----- 241  
373 LIAQVITLSTENVSKNIQIKIDNEDLNI SRANSSIKELKKDKKLLSTKEKELNSSFM 432  
242 -KLATEIAAIGIKETETTRPYVDYDMLSLKGAAKKM-NTCNEYQQRHGKKTILLE 299  
433 SKLANIEALGSGTINLSKVKSPTDYKQLGGGAAEAARGILLAYQLSVLQOIHGAKTCV 492  
300 VP 301  
493 PP 494

Query Match 7.7%; Score 116; DB 6; Length 941;  
Best Local Similarity 20.1%; Pred. No. 1.3;  
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7

7 EQTVGVKSAIETADGAL-DFYKYLDOVIP-WKTFTDETIKLSRFKQBYSOEASVVLGVD 64  
596 KEVVESINEATQNASAGLDFLATYDSEIVPNFNTAERTKMSKNTSQILKEADKKLPD 655  
65 IKVLLMDSQKQYFEATQTVYEWGVVTVQLLSAYILLFDEYNKKAQKDIILRLDDGV 124  
656 VKKLLBDSK-----GLV-----DGR 671  
125 NKLNEAQKSLGSSQSFNNSAGKLLALDSQ-----LTNDFSEKSSYFOSQVDIRK 175  
672 KKLADIKAEMPAETEKIKELADKIRDFESEEDLKIIRLLKNDVSKQDYFANPNLKEN 731  
176 EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELND---RLKAVQNF-----TSLSVT 228  
732 KLFAMENYGSAMSPFYTVLALWGALLVMSLLTVVHEBGANKYSHEIFYGRLLLFTLWG 791  
229 VKQA 232  
792 LSQA 795

Query Match 7.7%; Score 116; DB 6; Length 941;  
Best Local Similarity 20.1%; Pred. No. 1.3;  
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7

7 EQTVGVKSAIETADGAL-DFYKYLDOVIP-WKTFTDETIKLSRFKQBYSOEASVVLGVD 64  
596 KEVVESINEATQNASAGLDFLATYDSEIVPNFNTAERTKMSKNTSQILKEADKKLPD 655  
65 IKVLLMDSQKQYFEATQTVYEWGVVTVQLLSAYILLFDEYNKKAQKDIILRLDDGV 124  
656 VKKLLBDSK-----GLV-----DGR 671  
125 NKLNEAQKSLGSSQSFNNSAGKLLALDSQ-----LTNDFSEKSSYFOSQVDIRK 175  
672 KKLADIKAEMPAETEKIKELADKIRDFESEEDLKIIRLLKNDVSKQDYFANPNLKEN 731  
176 EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELND---RLKAVQNF-----TSLSVT 228  
732 KLFAMENYGSAMSPFYTVLALWGALLVMSLLTVVHEBGANKYSHEIFYGRLLLFTLWG 791  
229 VKQA 232  
792 LSQA 795

Query Match 7.7%; Score 116; DB 6; Length 941;  
Best Local Similarity 20.1%; Pred. No. 1.3;  
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7

7 EQTVGVKSAIETADGAL-DFYKYLDOVIP-WKTFTDETIKLSRFKQBYSOEASVVLGVD 64  
596 KEVVESINEATQNASAGLDFLATYDSEIVPNFNTAERTKMSKNTSQILKEADKKLPD 655  
65 IKVLLMDSQKQYFEATQTVYEWGVVTVQLLSAYILLFDEYNKKAQKDIILRLDDGV 124  
656 VKKLLBDSK-----GLV-----DGR 671  
125 NKLNEAQKSLGSSQSFNNSAGKLLALDSQ-----LTNDFSEKSSYFOSQVDIRK 175  
672 KKLADIKAEMPAETEKIKELADKIRDFESEEDLKIIRLLKNDVSKQDYFANPNLKEN 731  
176 EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELND---RLKAVQNF-----TSLSVT 228  
732 KLFAMENYGSAMSPFYTVLALWGALLVMSLLTVVHEBGANKYSHEIFYGRLLLFTLWG 791  
229 VKQA 232  
792 LSQA 795

Query Match 7.7%; Score 116; DB 6; Length 941;  
Best Local Similarity 20.1%; Pred. No. 1.3;  
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7

7 EQTVGVKSAIETADGAL-DFYKYLDOVIP-WKTFTDETIKLSRFKQBYSOEASVVLGVD 64  
596 KEVVESINEATQNASAGLDFLATYDSEIVPNFNTAERTKMSKNTSQILKEADKKLPD 655  
65 IKVLLMDSQKQYFEATQTVYEWGVVTVQLLSAYILLFDEYNKKAQKDIILRLDDGV 124  
656 VKKLLBDSK-----GLV-----DGR 671  
125 NKLNEAQKSLGSSQSFNNSAGKLLALDSQ-----LTNDFSEKSSYFOSQVDIRK 175  
672 KKLADIKAEMPAETEKIKELADKIRDFESEEDLKIIRLLKNDVSKQDYFANPNLKEN 731  
176 EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELND---RLKAVQNF-----TSLSVT 228  
732 KLFAMENYGSAMSPFYTVLALWGALLVMSLLTVVHEBGANKYSHEIFYGRLLLFTLWG 791  
229 VKQA 232  
792 LSQA 795

Query Match 7.7%; Score 116; DB 6; Length 941;  
Best Local Similarity 20.1%; Pred. No. 1.3;  
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7

7 EQTVGVKSAIETADGAL-DFYKYLDOVIP-WKTFTDETIKLSRFKQBYSOEASVVLGVD 64  
596 KEVVESINEATQNASAGLDFLATYDSEIVPNFNTAERTKMSKNTSQILKEADKKLPD 655  
65 IKVLLMDSQKQYFEATQTVYEWGVVTVQLLSAYILLFDEYNKKAQKDIILRLDDGV 124  
656 VKKLLBDSK-----GLV-----DGR 671  
125 NKLNEAQKSLGSSQSFNNSAGKLLALDSQ-----LTNDFSEKSSYFOSQVDIRK 175  
672 KKLADIKAEMPAETEKIKELADKIRDFESEEDLKIIRLLKNDVSKQDYFANPNLKEN 731  
176 EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELND---RLKAVQNF-----TSLSVT 228  
732 KLFAMENYGSAMSPFYTVLALWGALLVMSLLTVVHEBGANKYSHEIFYGRLLLFTLWG 791  
229 VKQA 232  
792 LSQA 795

Query Match 7.7%; Score 116; DB 6; Length 941;  
Best Local Similarity 20.1%; Pred. No. 1.3;  
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7

7 EQTVGVKSAIETADGAL-DFYKYLDOVIP-WKTFTDETIKLSRFKQBYSOEASVVLGVD 64  
596 KEVVESINEATQNASAGLDFLATYDSEIVPNFNTAERTKMSKNTSQILKEADKKLPD 655  
65 IKVLLMDSQKQYFEATQTVYEWGVVTVQLLSAYILLFDEYNKKAQKDIILRLDDGV 124  
656 VKKLLBDSK-----GLV-----DGR 671  
125 NKLNEAQKSLGSSQSFNNSAGKLLALDSQ-----LTNDFSEKSSYFOSQVDIRK 175  
672 KKLADIKAEMPAETEKIKELADKIRDFESEEDLKIIRLLKNDVSKQDYFANPNLKEN 731  
176 EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELND---RLKAVQNF-----TSLSVT 228  
732 KLFAMENYGSAMSPFYTVLALWGALLVMSLLTVVHEBGANKYSHEIFYGRLLLFTLWG 791  
229 VKQA 232  
792 LSQA 795

Query Match 7.7%; Score 116; DB 6; Length 941;  
Best Local Similarity 20.1%; Pred. No. 1.3;  
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7

7 EQTVGVKSAIETADGAL-DFYKYLDOVIP-WKTFTDETIKLSRFKQBYSOEASVVLGVD 64  
596 KEVVESINEATQNASAGLDFLATYDSEIVPNFNTAERTKMSKNTSQILKEADKKLPD 655  
65 IKVLLMDSQKQYFEATQTVYEWGVVTVQLLSAYILLFDEYNKKAQKDIILRLDDGV 124  
656 VKKLLBDSK-----GLV-----DGR 671  
125 NKLNEAQKSLGSSQSFNNSAGKLLALDSQ-----LTNDFSEKSSYFOSQVDIRK 175  
672 KKLADIKAEMPAETEKIKELADKIRDFESEEDLKIIRLLKNDVSKQDYFANPNLKEN 731  
176 EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELND---RLKAVQNF-----TSLSVT 228  
732 KLFAMENYGSAMSPFYTVLALWGALLVMSLLTVVHEBGANKYSHEIFYGRLLLFTLWG 791  
229 VKQA 232  
792 LSQA 795

Query Match 7.7%; Score 116; DB 6; Length 941;  
Best Local Similarity 20.1%; Pred. No. 1.3;  
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7

7 EQTVGVKSAIETADGAL-DFYKYLDOVIP-WKTFTDETIKLSRFKQBYSOEASVVLGVD 64  
596 KEVVESINEATQNASAGLDFLATYDSEIVPNFNTAERTKMSKNTSQILKEADKKLPD 655  
65 IKVLLMDSQKQYFEATQTVYEWGVVTVQLLSAYILLFDEYNKKAQKDIILRLDDGV 124  
656 VKKLLBDSK-----GLV-----DGR 671  
125 NKLNEAQKSLGSSQSFNNSAGKLLALDSQ-----LTNDFSEKSSYFOSQVDIRK 175  
672 KKLADIKAEMPAETEKIKELADKIRDFESEEDLKIIRLLKNDVSKQDYFANPNLKEN 731  
176 EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELND---RLKAVQNF-----TSLSVT 228  
732 KLFAMENYGSAMSPFYTVLALWGALLVMSLLTVVHEBGANKYSHEIFYGRLLLFTLWG 791  
229 VKQA 232  
792 LSQA 795

Query Match 7.7%; Score 116; DB 6; Length 941;  
Best Local Similarity 20.1%; Pred. No. 1.3;  
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7

7 EQTVGVKSAIETADGAL-DFYKYLDOVIP-WKTFTDETIKLSRFKQBYSOEASVVLGVD 64  
596 KEVVESINEATQNASAGLDFLATYDSEIVPNFNTAERTKMSKNTSQILKEADKKLPD 655  
65 IKVLLMDSQKQYFEATQTVYEWGVVTVQLLSAYILLFDEYNKKAQKDIILRLDDGV 124

DE N. gonorrhoeae amino acid sequence SEQ ID 180.  
 XX Antibacterial; infection; vaccine; gene therapy.  
 KW Neisseria gonorrhoeae.  
 XX WO200279243-A2.  
 XX 10-OCT-2002.  
 XX 12-FEB-2002; 2002WO-IB002069.  
 XX 12-FEB-2001; 2001GB-00003424.  
 XX (CHIR-) CHIRON SPA.  
 XX Fontana MR, Pizza M, Masignani V, Monaci E;  
 DR WPI; 2003-058415/05.  
 DR N-PSDB; ABZ37795.  
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection.  
 XX Claim 1; Page 197; 815pp; English.  
 XX The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention  
 XX Sequence 1015 AA;  
 SQ

Query Match 7.5%; Score 113; DB 6; Length 1015;  
 Best Local Similarity 21.8%; Pred. No. 2.5;  
 Matches 75; Conservative 54; Mismatches 113; Indels 102; Gaps 17;

QY 9 TVEVVKSAIETADGALDFYKYLQVDPWK-----FDETIKELSRPKQBYSQBASVLVG 63  
 DB 104 TADFAKSLVLTAD-AMQINSQVRQVTSSETEYLAQQQLDNTANRTRASLESTANLYVS 162  
 QY 64 DIKVL--MDSQD--KYFEATQTVYEWCGV-----VTOLLSAYILLFDEYNEK 107  
 DB 163 TGRALKDYGVTQOEILKFTFAANNAMTIGVGAAQQAALMQLSQALGSGVLQGEF--K 220  
 QY 108 KASAKDILIRIL-----DDGVNKL-NEAQ-----KSLGSSQSFNNASGKLLALD 152  
 DB 221 SISEAAPILLDTIAEYMGKSRDEIKLGSEGLTADVIFKAISGASEKFGEQAARMPVTM 280  
 QY 153 SQTNDPSEKSSYFQSOVDRIKE-----AYAGAAA- 183  
 DB 281 GOALTTFVSN--WQSVKSLNDSGMTSGIAAVIKLIADNLNLVPIVAGFAVAVA 337  
 QY 184 -----GIVAGPGLIISYTAAGVIRGKLIPELNDRLKAVQNFPTSLSVTVKQAKND 235  
 DB 338 APTIALNLALLANPFGII---AVAIGTVVG-LIAKFGDEIDVFGGWSNLS-----D 385  
 QY 236 IDAAKLKLAETIA--AIGEIKTETETTFYVDYDMLSLKGA 277  
 DB 386 VIRAVQVIITETVGEAVGTVKS-----FDGLTGRINEGA 420

RESULT 8  
 ABP78617  
 ID ABP78617 standard; protein; 1015 AA.  
 XX  
 AC ABP78617;  
 XX 07-MAR-2003 (first entry)  
 DT

XX N. gonorrhoeae amino acid sequence SEQ ID 3764.  
 DE Antibacterial; infection; vaccine; gene therapy.  
 KW Neisseria gonorrhoeae.  
 XX WO200279243-A2.  
 XX 10-OCT-2002.  
 XX 12-FEB-2002; 2002WO-IB002069.  
 XX 12-FEB-2001; 2001GB-00003424.  
 XX (CHIR-) CHIRON SPA.  
 XX Fontana MR, Pizza M, Masignani V, Monaci E;  
 DR WPI; 2003-058415/05.  
 DR N-PSDB; ABZ39587.  
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection.  
 XX Disclosure; Page 460; 815pp; English.  
 XX The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention  
 XX Sequence 1015 AA;  
 SQ

Query Match 7.5%; Score 113; DB 6; Length 1015;  
 Best Local Similarity 21.8%; Pred. No. 2.5;  
 Matches 75; Conservative 54; Mismatches 113; Indels 102; Gaps 17;

QY 9 TVEVVKSAIETADGALDFYKYLQVDPWK-----FDETIKELSRPKQBYSQBASVLVG 63  
 DB 104 TADFAKSLVLTAD-AMQINSQVRQVTSSETEYLAQQQLDNTANRTRASLESTANLYVS 162  
 QY 64 DIKVL--MDSQD--KYFEATQTVYEWCGV-----VTOLLSAYILLFDEYNEK 107  
 DB 163 TGRALKDYGVTQOEILKFTFAANNAMTIGVGAAQQAALMQLSQALGSGVLQGEF--K 220  
 QY 108 KASAKDILIRIL-----DDGVNKL-NEAQ-----KSLGSSQSFNNASGKLLALD 152  
 DB 221 SISEAAPILLDTIAEYMGKSRDEIKLGSEGLTADVIFKAISGASEKFGEQAARMPVTM 280  
 QY 153 SQTNDPSEKSSYFQSOVDRIKE-----AYAGAAA- 183  
 DB 281 GOALTTFVSN--WQSVKSLNDSGMTSGIAAVIKLIADNLNLVPIVAGFAVAVA 337  
 QY 184 -----GIVAGPGLIISYTAAGVIRGKLIPELNDRLKAVQNFPTSLSVTVKQAKND 235  
 DB 338 APTIALNLALLANPFGII---AVAIGTVVG-LIAKFGDEIDVFGGWSNLS-----D 385  
 QY 236 IDAAKLKLAETIA--AIGEIKTETETTFYVDYDMLSLKGA 277  
 DB 386 VIRAVQVIITETVGEAVGTVKS-----FDGLTGRINEGA 420

RESULT 9  
 ABP07451  
 ID ABP07451 standard; protein; 1036 AA.  
 XX  
 AC ABP07451;  
 XX

DT 12-FEB-2004 (first entry)  
DE Bacterial polypeptide #3564.  
KW Proteus mirabilis infection; bacterial infection; antibacterial;  
XX immunostimulant.  
XX Proteus mirabilis.  
OS US6605709-B1.  
PN 12-AUG-2003.  
XX 05-APR-2000; 2000US-00543681.  
PF 09-APR-1999; 99US-0128706P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PA Breton GL;  
XX WPI; 2003-895291/82.  
PI N-PSDB; ADF03279.  
XX  
DR New Proteus mirabilis polypeptides and polynucleotides, useful as  
XX reagents for diagnosis of bacterial disease, as components of  
PT antibacterial vaccines, as targets for antibacterial drugs, or as  
PT biocontrol agents for plants.  
XX  
PS Disclosure; SEQ ID NO 7736; 870pp; English.  
XX  
CC The invention relates to new Proteus mirabilis polypeptides and  
CC polynucleotides. The invention also relates to antibodies against the  
CC polypeptides. Methods for producing the polypeptides, a method of  
CC generating vaccines for immunising an individual against P. mirabilis, a  
CC method for evaluating a compound for the ability to bind a P. mirabilis, a  
CC polypeptide and a method for screening test compounds for anti-bacterial  
CC activity. The polypeptides and polynucleotides are useful as molecular  
CC targets for diagnosing, preventing and treating pathological conditions  
CC resulting from bacterial infection, as reagents for diagnosis of  
CC bacterial diseases, as components of antibacterial vaccines, as targets  
CC for antibacterial drugs or as bio-control agents for plants. This  
CC sequence represents a Proteus mirabilis polypeptide of the invention.  
XX  
SQ Sequence 1036 AA;  
  
Query Match 7.4%; Score 111.5; DB 7; Length 1036;  
Best Local Similarity 21.0%; Pred. No. 3.4;  
Matches 69; Conservative 62; Mismatches 117; Indels 81; Gaps 14;  
  
Qy 2 TGIFAEQTVVVK-----SAITADGALDFVYKYLDPVWKTFTDETIKELSRKQE 53  
Db 167 TALSGMEINELIKQKNGENVSSSELAKASIDLINQLVD-----TASSLNNNISAFSQ 220  
Qy 54 YQSEASVL-----VGDIKVLLMDSQDKYFEATQTVYEWGCVVTTQLLSAYILL-PDEY 104  
Db 221 LNKGLSVLSNTKHLGVGN-KLQNLPLNDKLGTLGDTV---SGILSAISASFILSNADAD 276  
Qy 105 NEKASAKDILIRLDD---GVNKLNEAOKSLIGSSOSFNNAAGKLLA-----LD 152  
Db 277 TGTAAAGVETTTKVLGNVGAQVSIYLAQVLAQGLSTS---AASAGLIASAVTLAISPLS 334  
Qy 153 SOLTNDPSEKSYFQSDVRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELN 212  
Db 335 FLAIADQPKRANKIEYSQRFKKFYEGDS-----LIAAFKFTGALDASL----- 380  
Qy 213 DRLKAVQNFFTSLSVTVVQKANKIDAALK--LATEIAAI-----GEIKTETETTRFYVD 265  
Db 381 -----TTINTALGTISAGISAASASLIGAPISALVGAIGTIGSILEASK----- 426  
Qy 266 YDDLMLSLKGAOKMINTCNEYQORHOK 294  
Db 427 -----QSMFHVANRMANTIAEWKTHGK 450

RESULT 10  
ABJ19025  
ID ABJ19025 standard; protein; 2681 AA.  
XX  
AC ABJ19025;  
XX  
DT 06-MAR-2003 (first entry)  
XX  
DE Pathogen specific antigen related staphylococcal protein SEQ ID No 274.  
XX  
KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;  
KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
KW autoimmune disease; HIV; hepatitis.  
XX  
OS Staphylococcus sp.  
XX  
XX WO200259148-A2;  
XX  
XX 01-AUG-2002.  
XX  
XX 21-JAN-2002; 2002WO-EP000546.  
XX  
XX 26-JAN-2001; 2001AT-00000130.  
XX  
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.  
XX  
XX Meinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W;  
PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;  
PI Tempelmeier B;  
XX  
XX WPI; 2003-075410/07.  
XX  
XX Identifying, isolating and producing hyperimmune serum-reactive antigens  
XX from a pathogen, for preparing vaccine or medicament for treating or  
XX preventing e.g. staphylococcal infections, comprises providing antibody  
XX preparation.  
XX  
XX Example 7; Page 200; 252pp; English.  
XX  
CC The invention relates to a novel method for identifying, isolating and  
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
CC allergen, a tissue or host prone to auto-immunity, where the antigens are  
CC used in a vaccine, comprises providing antibody preparation from a plasma  
CC pool of a type of animal, or individual sera with antibodies against the  
CC specific pathogen, tumour, allergen, tissue or host prone to auto-  
CC immunity. The hyperimmune serum-reactive antigens comprising any of the  
CC 62 sequences of 53-2261 amino acids fully defined in the specification,  
CC or their hyperimmune fragments are useful for the manufacture of a  
CC pharmaceutical preparation, particularly a vaccine against staphylococcal  
CC infections or colonisation against S. aureus or S. epidermidis. The  
CC preparation of antibodies is useful for the manufacture of a medicament  
CC for treating or preventing staphylococcal infections or colonisation  
CC against S. aureus or S. epidermidis. The antibody preparations may also  
CC be used for diagnostic and imaging purposes. Other conditions that can be  
CC treated include cancer, autoimmune diseases or infections caused by viral  
CC (e.g. HIV, hepatitis A, B or C), fungal or protozoan pathogens. This  
CC sequence represents a staphylococcal protein relating to the method for  
CC identifying and producing pathogen specific antigens of the invention  
XX  
SQ Sequence 2681 AA;  
  
Query Match 7.3%; Score 110.5; DB 6; Length 2681;  
Best Local Similarity 20.7%; Pred. No. 14;  
Matches 71; Conservative 52; Mismatches 129; Indels 91; Gaps 13;  
  
Qy 2 TGIFAEQTVVVKSAIETADGALDFVYKYLDPVWKTFTDET-----IKELSR 49  
Db 998 TDLKSEKLDLTKS-YKLTDTSFNDVFKYIDEMTMDRIADTEKVNKKYKTYLQKNLSD 1056  
Qy 50 FKQBY-----SQEASVLVGDIKVL-----LMSDQDKYFEATQTVYEW--G 88



Db 1057 YMKYDNCILEISKYNDAAKVLGDFTATATELQNDQFQVQKDNWAEFKQITLESFKDG 1116  
 Qy 89 VVTQLLSAYILLDFDYNEKKASAKQDIILIRLLDGVNKNLEAKQKSLGSSQSFNNASKL 148  
 Db 1117 IVT-----BAEKARLRVQLDMLDRESMD-----IEERYKSLLANQVTTNTDIKNRL 1161  
 Qy 149 LALDSQLTDFSEKSYFSQSVDRTRKEAYAGAAAGIVAGPGLIISYSIAAGVLEGKLI 208  
 Db 1162 TA-----SRPFLSVHAS-LRK-----VIEQIIADGRVDESEK 1193  
 Qy 209 PELNRLKAVONFFTSLSVTYKQANKDI-----DAAKLKLATEIAAIGEIKTETETTRF 262  
 Db 1194 TLANNLSLNTYNTTLAYSKTIQEAALNTLSQIITSSDVASKKVEEFGVITTISSDVTIKK 1253  
 Qy 263 YVD-----YDDLMLSLKGAKKMINTCNEYQQRHGKKTILLE 299  
 Db 1254 QRDGAVITYYYSVGVPTLSNDPAKSW--TTNDLKOLHIKDMYLD 1294

## RESULT 11

ABU42996

ID ABU42996 standard; protein; 1189 AA.

AC ABU42996;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #28523.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Staphylococcus epidermidis.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX N-PSDB; ACN46866.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 70920; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX proliferation or the activity of a gene in an operon required for

XX proliferation; (7) identifying a compound that influences the activity of

XX the gene product or that has an activity against a biological pathway

XX required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 1189 AA;

## Query Match

7.3%; Score 110; DB 6; Length 1189;

Best Local Similarity 18.3%; Pred. No. 5.4; Indels 82; Gaps 9;

Matches 61; Conservative 60; Mismatches 130; Indels 82; Gaps 9;

Qy 7 EQTVVVK-----SAIETADGALDFYKYLQDVIPWKTDFETIKELSRFKQYSQASV 60

Db 163 EESAGVLKYKKAESIQKLDHTEDNLNRVEDILYDLGRVPELKEAAIAKEYKQLSKE 222

Qy 61 LVGDIKVLMDSQKYFEATQTVYEW-----GVVTLLSAILLFDEYNE 106

Db 223 MEQSDVITVSDIDHYTENDQRLDERLNHLKSQAQKEGQQAQINQLQRY----- 273

Qy 107 KKAQAQDILIRLLDGVNKNLEAKQKSLGSSQSFNNASGLLALDLSQLTND-----FS 160

Db 274 -GKRQON-----DYDIEKLN---YELVATENYQLSGKLVLEERKQKQSTNARYE 323

Qy 161 EKSYFQSVDRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNDRL----K 216

Db 324 EELNLESQDISIKNEKAQNE-----KLLADLKNQKQKQLNK 359

Qy 217 AVQNFSTLSVTYVQANKDDAAKLAETAAIGETETETTRFYVDYDMLMLSLKG 276

Db 360 EVQELSLIYVSDQHDQKLEEKINSYITLMSEQSDVNNDIRFLEHTINENEAKKSRLDS 419

Qy 277 AAKMINTCNEYQO-----BHGKKTLLLEV 300

Db 420 RLVEAFNLQDKIQNITQTNKAYQSSKKSMQKV 452

## RESULT 12

ADC95690

ID ADC95690 standard; protein; 515 AA.

XX AC ADC95690;

XX 01-JAN-2004 (first entry)

XX E. faecium protein sequence SEQ ID 5317.

XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;

XX abdominal-pelvic infection.

XX Enterococcus faecium.

XX US583275-B1.

XX 24-JUN-2003.

XX 30-JUN-1998; 98US-00107532.

XX 02-JUL-1997; 97US-0051571P.

XX 14-MAY-1998; 98US-0085598P.





Db 325 OTSLAEAK-----KVK-----HPEKLOESLMKAAAEKETIQ 357

Oy 281 -----WINTCNEYQ--QRHGKKTLLLEVDP 302

Db 358 KAAANLIKTQOELEKYQKSTKELLALRD 386

RESULT 14

ABU25599

ID ABU25599 standard; protein; 742 AA.

XX AC ABU25599;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #1126.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Clostridium difficile.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR N-PSDB; ACA29469.

XX DR WPI: 2003-029926/02.

XX DR N-PSDB; ACA29469.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 53523; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC polypeptide or its fragment whose expression is inhibited by the

XX CC antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC proliferation or the activity of a gene in an operon required for

XX CC proliferation; (7) identifying a compound that influences the activity of

XX CC the gene product or that has an activity against a biological pathway

XX CC required for proliferation, or that inhibits cellular proliferation; (8)

XX CC identifying a gene required for cellular proliferation or the biological

XX CC pathway in which a proliferation-required gene or its gene product lies

XX CC or a gene on which the test compound that inhibits proliferation of an

XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX CC compound's activity; (11) a culture comprising strains in which the gene

XX CC product is overexpressed or underexpressed; (12) determining the extent

XX CC to which each of the strains is present in a culture or collection of

XX CC strains; or (13) identifying the target of a compound that inhibits the

XX CC proliferation of an organism. The antisense nucleic acids are useful for

XX CC identifying proteins or screening for homologous nucleic acids required

XX CC for cellular proliferation to isolate candidate molecules for rational

XX CC drug discovery programs, or for screening homologous nucleic acids

XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 742 AA;

Query Match 7.1%; Score 107.5; DB 6; Length 742;

Best Local Similarity 21.2%; Pred. No. 4.7; Mismatches 63; Gaps 10;

Matches 66; Conservative 49; Indels 134; Indels 63; Gaps 10;

Oy 18 ETADGALDFYKYLQVIVPWTETIKELSRKQEVSEASVLVGDIKVLMDSQDYF 77

Db 236 KTVDLASDATSKLSDTVKDIKSDLPITIKT-----LNDTKLLSSDLKKFLEDNDNL 288

Oy 78 BATQTVVWCGVTVQLLSAYILLFDEYNEKKAQKQILIRLD-----DGVNKLNEAQ 131

Db 289 ELSPLIKSDLMVLDL-----SSASSLTLLIDAVNSGSEDPVKLIDNL 333

Oy 132 KSLIGSSQSFNNASGKLLALDSOLTN-----DFSEKSSYFQSDVDRIRKEAYA 179

Db 334 SEKLSNLQSLNDTLVDFLTQKLNQLTNNRLDDVDINLEDSSNKIDSSISTINDIKNKVIS 393

Oy 180 GAAAGIVAGPFGLIISYIAAGVIEGKLIPELNDRL-KAVQNFFTSLSVTVKQANKDI-- 236

Db 394 GQPSISA--LNNVLSLSNGIGRINLNLNFDKSKIPINNIFAN-----SIKVANDIITV 448

Oy 237 -DAAKLK-----LATIAAIGEIKTETETRTFYVDYDDLMLSLLKGAAKKMINTCNEY 288

Db 449 LDKAEAKLPKVEEILTTSKLKSGNAQESISLIR-----BRLPLAKGMLDLDLITLSKI 501

Oy 289 QQRHGKKTLLLEV 300

Db 502 SNGEDMKKLVSL 513

RESULT 15

ABU43311

ID ABU43311 standard; protein; 3692 AA.

XX AC ABU43311;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #28838.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Staphylococcus epidermidis.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI: 2003-029926/02.

XX DR N-PSDB; ACA47181.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 71235; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SEQ Sequence 3692 AA;

Query Match 7.1%; Score 106.5; DB 6; Length 3692;

Best Local Similarity 21.5%; Pred. No. 46;

Matches 71; Conservative 49; Mismatches 129; Indels 81; Gaps 12;

Qy 11 EVKSAIETADGALDFYKYLDDQVTPKTPDETTELKLSR-----FKQYSQ-EA 58  
 Db 564 EQVNDIIPSNYTLASYNKY-----NKLKERAQTVDLBETNTPNQRYSQTQI 612  
 Qy 59 SVLVGDIKVLIM-----DSQDKYFEATQVYEWCGVVTQLLSAYILLDFEYNEKKASA 111  
 Db 613 DLLLHELOTTLINRVSAISREINDKAQEMTDAVYDSTELTTEKDT---LVDQIENHNEI 669  
 Qy 112 QKILIRILDGQVKNLNEAKSLGSS-----QSFNNSAGKLLALDSQLTNDFS 160  
 Db 670 SNNIDDELTDGVERVKEAGLHTLESPTPHVTPKPNARQVNNRA-----DQCKTLIRN 723  
 Qy 161 EKSSYFQSDVRIRK-EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELNDRLKAVQ 219  
 Db 724 NHEATTEQNEAIRQVEAHSDDA-----IAKIGEAETDTTVNE---ARD 764  
 Qy 220 NFFTSLSVTVKQANKDIDAAKILKATETAIAIGETETETTRFVVDVDDMLSLKGAAK 279  
 Db 765 NGTKLIATDVNPPTKKAEEA---RAAVTNSANSKIKDINNNTQATILDERDAIALVNRSD 821  
 Qy 280 KMINTCNEYQ-----QRHGKKTLLFVP 301  
 Db 822 EAIQNINTAQGNDVDVTEAQNGTNTIQQVP 851

RESULT 16

ABP40235

XX ID ABP40235 standard; protein; 3696 AA.

XX AC ABP40235;

XX

DT 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5080.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 XX antibacterial; gene therapy.

OS Staphylococcus epidermidis.

PN US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-00134001.

XX 14-AUG-1997; 97US-0055779P.

PR 08-NOV-1997; 97US-0064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI: 2002-381255/41.

XX N-PSDB; ABN92780.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermis  
 PT polypeptide, useful for diagnosing and treating bacterial infections.

XX Disclosure; SEQ ID NO 5080; 267pp; English.

XX ABN90338 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly *S. epidermidis* infections. The sequences can be used to  
 CC screen for compounds able to interfere with the *S. epidermidis* life cycle  
 CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site

XX SEQ Sequence 3696 AA;

Query Match 7.1%; Score 106.5; DB 5; Length 3696;

Best Local Similarity 21.5%; Pred. No. 47;

Matches 71; Conservative 49; Mismatches 129; Indels 81; Gaps 12;

Qy 11 EVKSAIETADGALDFYKYLDDQVTPKTPDETTELKLSR-----FKQYSQ-EA 58  
 Db 568 EQVNDIIPSNYTLASYNKY-----NKLKERAQTVDLBETNTPNQRYSQTQI 616  
 Qy 59 SVLVGDIKVLIM-----DSQDKYFEATQVYEWCGVVTQLLSAYILLDFEYNEKKASA 111  
 Db 617 DLLLHELOTTLINRVSAISREINDKAQEMTDAVYDSTELTTEKDT---LVDQIENHNEI 673  
 Qy 112 QKILIRILDGQVKNLNEAKSLGSS-----QSFNNSAGKLLALDSQLTNDFS 160  
 Db 674 SNNIDDELTDGVERVKEAGLHTLESPTPHVTPKPNARQVNNRA-----DQCKTLIRN 727  
 Qy 161 EKSSYFQSDVRIRK-EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELNDRLKAVQ 219  
 Db 728 NHEATTEQNEAIRQVEAHSDDA-----IAKIGEAETDTTVNE---ARD 768  
 Qy 220 NFFTSLSVTVKQANKDIDAAKILKATETAIAIGETETETTRFVVDVDDMLSLKGAAK 279  
 Db 769 NGTKLIATDVNPPTKKAEEA---RAAVTNSANSKIKDINNNTQATILDERDAIALVNRSD 825  
 Qy 280 KMINTCNEYQ-----QRHGKKTLLFVP 301  
 Db 826 EAIQNINTAQGNDVDVTEAQNGTNTIQQVP 855

RESULT 17



CC injury (CCI) and spared nerve injury (SNI) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3187 AA;

Query Match 7.0%; Score 106; DB 7; Length 3187;  
Best Local Similarity 22.2%; Pred. No. 42;  
Matches 78; Conservative 53; Mismatches 134; Indels 86; Gaps 16;  
Qy 6 AEQTV--EVVKSATETADGALDFYNYLDQVVPKTFDETIKELSRFKQYQSQASVILVG 63  
Db 1748 AEATLLANSAPGVSETFSSHDDINNYLQDLK--GRIAELEMEKQKDRLSQTLN 1804  
Qy 64 DIKVL-----MDSQKYFE-----ATQTVYEWCGVVTQLLSAYILLDFEYNEKAS 110  
Db 1805 EKNALLTOISAKDSKLEBEVAKINMLNQIOEELSRVTKL-----KETAE 1852  
Qy 111 AQKDILIRILDDGVNKLNEAKSLGSSQSFNNASGKLALDSOLTN-----DFSEKSS 164  
Db 1853 EEKDDLBERL---WNQIAELNGSIGNYYQVTDQIKNEQLESEMQLKRCVSELEEKQ 1909  
Qy 165 YF---QSQVD--RIRKEAYAGAAAGIVAGPFGLLIISYSTAAGVIEGKLIPELNDRLKAVQ- 219  
Db 1910 QLVKEKTKVESEIRKE-YMEKIQAQKGPQSGKIHAKEIQ-----ELLKEKQEQVKQLQK 1962  
Qy 220 -----NFTSLSVTVK-----QANKDIDAAKKLATEI-----AATGEIK----- 254  
Db 1963 DCIRYLGRISALEKTVKALEFVHTESQKDLDTGNLAQAVEHHKKAQAEELSSFKILLDD 2022  
Qy 255 TETETTFRYVD-----YDDLMLSLKGAAKMINTCNEYQORHGKK 295  
Db 2023 TQSEARVADNLNKLKELQSNKESIKSQIKQKDEDLRLRLEQAEKIRKE 2073

RESULT 19  
ID ADE56035 standard; protein; 3187 AA.  
XX ADE56035;  
AC ADE56035;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Rat Protein BAA05026, SEQ ID NO 1874.  
XX  
KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
XX  
OS Rattus norvegicus.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GENO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
DR WPI; 2003-268312/26.  
DR GENBANK; BAA05026.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3187 AA;

Query Match 7.0%; Score 106; DB 7; Length 3187;  
Best Local Similarity 22.2%; Pred. No. 42;  
Matches 78; Conservative 53; Mismatches 134; Indels 86; Gaps 16;  
Qy 6 AEQTV--EVVKSATETADGALDFYNYLDQVVPKTFDETIKELSRFKQYQSQASVILVG 63  
Db 1748 AEATLLANSAPGVSETFSSHDDINNYLQDLK--GRIAELEMEKQKDRLSQTLN 1804  
Qy 64 DIKVL-----MDSQKYFE-----ATQTVYEWCGVVTQLLSAYILLDFEYNEKAS 110  
Db 1805 EKNALLTOISAKDSKLEBEVAKINMLNQIOEELSRVTKL-----KETAE 1852  
Qy 111 AQKDILIRILDDGVNKLNEAKSLGSSQSFNNASGKLALDSOLTN-----DFSEKSS 164  
Db 1853 EEKDDLBERL---WNQIAELNGSIGNYYQVTDQIKNEQLESEMQLKRCVSELEEKQ 1909  
Qy 165 YF---QSQVD--RIRKEAYAGAAAGIVAGPFGLLIISYSTAAGVIEGKLIPELNDRLKAVQ- 219  
Db 1910 QLVKEKTKVESEIRKE-YMEKIQAQKGPQSGKIHAKEIQ-----ELLKEKQEQVKQLQK 1962  
Qy 220 -----NFTSLSVTVK-----QANKDIDAAKKLATEI-----AATGEIK----- 254  
Db 1963 DCIRYLGRISALEKTVKALEFVHTESQKDLDTGNLAQAVEHHKKAQAEELSSFKILLDD 2022  
Qy 255 TETETTFRYVD-----YDDLMLSLKGAAKMINTCNEYQORHGKK 295  
Db 2023 TQSEARVADNLNKLKELQSNKESIKSQIKQKDEDLRLRLEQAEKIRKE 2073  
RESULT 20  
ID ABB71280 standard; protein; 588 AA.  
XX ABB71280;  
AC ABB71280;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 40632.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL15383.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.

XX Disclosure; SEQ ID NO 40632; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signaling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 588 AA;

Query Match 7.0%; Score 105.5; DB 4; Length 588;

Best Local Similarity 20.5%; Pred. No. 5.1;

Matches 66; Conservative 44; Mismatches 87; Indels 125; Gaps 13;

QY 11 EVVKGAIETADGALDFYNYKLDQV---IPWK----- 38

DB 300 EILALLAENAGALDRFNTQLDVGAINWKIDAEILDNQAEYLDATVTEELLRNQTBELL 359

QY 39 -----TFD-----ETIKELSRFKQEQVQASVLVGDIKVLLMDSQDKY 76

DB 360 LWEVELLRGVVTSFQNLDFANRSIEAVSDTLRLQEQNKDRVRLNVLVDKLVN---TQGOI 415

QY 77 FEATQTVYEWCGVVTOLLISAYILLPDEYNEKKASAKQDILIRLDDGVNKLNEAQ----- 131

DB 416 LRRTKGLDRLNLFVQLLGYI-----EPKNS-----LSDSPNLMKNSQINSILI 460

QY 132 --KSLIGSQSFNNASGKLLALDSQTLNDFSKSSYFQSDVRIRKEAYAGAAAGIVAGP 189

DB 461 ELKNVPEVRNLTKTSIRKLSFLDNQL-----ALFNQIQENRY----- 497

QY 190 FGLIISYSTAAGVIEGKLIPENLDRKAVQNFSTLSVTVKQANKDIDAKKLATEIAA 249

DB 498 -----YSVEAVI---KAWTPTN--LKEINDLTHALSISQKRTD-----LAIAISG 537

QY 250 IGEIKTETETTRF--VVDYDDL 269

DB 538 SAEYNTETPTPTFISYKGIEDI 559

RESULT 21

ID ABB77986

XX ABB77986 standard; protein; 1292 AA.

XX AC ABB77986;

XX 22-OCT-2002 (first entry)

XX Amino acid sequence of an Arabidopsis RAD50 homologue.

XX Nucleic acid integration; homologous recombination; telomeric region; RAD50.

XX Arabidopsis thaliana.

XX EPI217074-A1.

XX 26-JUN-2002.

XX 22-DEC-2000; 2000EP-00204693.

XX 22-DEC-2000; 2000EP-00204693.

XX (UYLE-) RIJKSUNIV LEIDEN.

XX (BINA-) STICHTING BINAIR VECTOR SYSTEEM.

XX Hooykaas PJJ, Van Attikum H, Bundock P;

XX WPI; 2002-550409/59.

XX Directing integration of nucleic acid of interest to a sub-telomeric region in an eukaryote with preference for non-homologous recombination, by steering an integration pathway towards homologous recombination.

XX Disclosure; Fig 5; 63pp; English.

The specification describes a method for directing integration of a nucleic acid of interest to a pre-determined site, where the nucleic acid has homology at or around the pre-determined site, in a eukaryote with a preference for non-homologous recombination. The method comprises steering an integration pathway towards homologous recombination. The method is useful for directing integration of a nucleic acid of interest to a subtelomeric and/or telomeric region in an eukaryote with a preference for non-homologous recombination. The nucleic acid of interest comprises an inactive gene to replace an active gene, or vice versa, is a portion of a gene delivery vehicle, confers a desired property to the eukaryotic cell, or encodes a therapeutic proteinaceous substance or a substance conferring resistance for an antibiotic substance to a cell. The method is useful for improving gene targeting efficiency. The method is useful in the replacement of an active gene by an inactive gene, for e.g. for the inactivation of genes controlling undesired side branches of metabolic pathways, to increase the quality of bulk products such as starch, or to increase the production of specific secondary metabolites or to inhibit formation of unwanted metabolites, and also to inactivate genes controlling senescence in fruits and flowers or that determine flower pigments. The method is also useful for replacing an inactive gene by an active gene. For e.g. the replacement of a defective p53 by an intact p53. Many tumours acquire a mutation in p53 during their development which renders it inactive and often correlates with a poor response to cancer therapy. By replacing the defect p53 by an intact p53, e.g. through gene therapy, conventional anti cancer therapy have better changes of succeeding. The method is also useful for therapeutic proteinaceous substance integration. A tumoricidal gene can be delivered to a pre-determined site present only in e.g. proliferating cells, or present only in tumour cells, e.g. to the site where a tumour antigen is expressed form. ABB77984-86 represent RAD50 homologues. RAD50 is involved in non-homologous recombination

XX Sequence 1292 AA;

Query Match 6.9%; Score 104.5; DB 5; Length 1292;

Best Local Similarity 15.0%; Pred. No. 17;

Matches 64; Conservative 93; Mismatches 133; Indels 137; Gaps 13;

QY 7 EOTVEVKSALJETADGALDFYNYKLDQV-----IPKTFDETIKE 46

DB 246 ETSVQKDAEVHNKEMMLKDLRLQDQVSIKTAERSTLFKEQQRQYALPEEN-EDTIEE 304





antibiotics, comprise sequences of antisense nucleic acids.

Example 3; SEQ ID NO 10796; 51lpp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 722 AA;

Query Match 6.9%; Score 104; DB 4; Length 722;

Best Local Similarity 19.8%; Pred. No. 8.9; Mismatches 109; Indels 108; Gaps 12; Matches 66; Conservative 51;

16 ALETADGALDFYKYLQDVIPWKTFTETIKELSRFQY-----SQE 57  
122 AVKGNKDKISYDDLKGTVAAKVGTSANFLEKNKEKYDYTIKNFDDATGLYKALENGE 181  
58 ASVLVGDIKVL-----LMSQD-----KYFEATQTVVWCG 88  
182 ADAIVDDYFVLGVAVXNGKQLGVGKETGSSYGFVAVKQGNPELIKFNAGLKNLKONG 241  
89 VVTLSSAVILLFDEYNEKKAQAQKILIRILDDGVNKLNEAQAQKLLGSSQS-----FN 142  
242 TYDKILNNVLTGDEINTQAGEQ-----MKITPKKEKYVIASDSTFAPFBEQ 290  
143 NASGKLLALDSQLTNDFSEKSY-----FQSQVDRIRKAYAGAAAGIVAGPFGLII 194  
291 NAQGDYVIGIDVILVKRAELQGFTEFKFIFGFSVAQAVE-----SQADGMVAG----- 340  
195 SYSIAAGVIEGKLIPELNDRLKA-----VQNFTSLSVTVKQAN---KQIDAQK-----LK 242  
341 -----WTITDRKKAQAFDSVPYFDSGIGIAVKGNDKIKSYDDLKGGKVGK 387  
243 LATEIA-AIGEIKTETETTRFFVYDDLMMLSLK 275  
388 IGTESADFLKKNKKYDYSIKYLDTTDALYSALE 421

RESULT 24

ADA32997  
ID ADA32997 standard; protein; 815 AA.

XX ADA32997;

XX 20-NOV-2003 (first entry)

XX Acinetobacter baumannii protein #158.

XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;  
XX plant biocontrol agent.

XX Acinetobacter baumannii.

XX US6562958-B1.

XX 13-MAY-2003.

XX

04-JUN-1999; 99US-00328352.

09-JUN-1998; 98US-0088701P.

(GENO-) GENOME THERAPEUTICS CORP.

Breton G, Bush D;

WPI; 2003-576092/54.

N-PSDB; ADA28871.

New Acinetobacter baumannii proteins and nucleic acids, useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.

Example; SEQ ID NO 4284; 328pp; English.

The invention relates to isolated Acinetobacter baumannii nucleic acids. The A. baumannii nucleic acids and polypeptides are useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, to detect the presence of A. baumannii and other Acinetobacter species in a sample, in screening compounds for the ability to interfere with the A. baumannii life cycle or to inhibit A. baumannii infection, and as biocontrol agents for plants. The present sequence represents the amino acid sequence of an A. baumannii protein.

Sequence 815 AA;

Query Match 6.9%; Score 104; DB 6; Length 815;

Best Local Similarity 22.0%; Pred. No. 10; Mismatches 123; Indels 86; Gaps 15; Matches 72; Conservative 47;

9 TVEVKSIAETADGALDFYKYLQDVIPWKTFTETIKELSRFQK--EYSQFASVLVGDIK 66  
220 TAHIIQSWADKADTEIIIPQITILEFGLPYFDEVIKEAESFKPEAKDEGRVLDRLA 279  
67 VLLMDSQKFEATQTVVWCGVTVQLLSAYILLFDE--YNEKKAQAQKILIRI----- 119  
280 LVITDGEDARD-----FDDAVYAEKRPGGGYRVVVAIADVSH 316  
120 ---LDDGVNKLNEAQAQKLLGSSQSFNNASGKLLALDSQLTNDFSEKSYFQSQVDRI--- 173  
317 YVRLDSALNE--EAER--GTSVYFPHF--VLPMLPEALSNGLCSLNPH---VDRLCW 366  
174 --RKEAYAGAAAGIVAGPFGL-----IISYSIAAGVIEG--KLIPELNDRLKAVQNFTSL 225  
367 CDLXLSRTGRVTGYEFYPAVMHKSARLTYYQVGQYFEGATDAIPKORDIHKSLNTLF-QL 425  
226 SVTVKQAKNDIDAQKLLKLAETIAAIGEIKTETETTRFFVYDDLL-----MLSLKGAQK 280  
426 YQILKNLVRDRA-----MEFTIETMTFDELGCIKEILPRTENDAHK 469  
281 MINTC-----NEYQORHGKKTILEV 300  
470 LIECMLLANVAAYALEHDIPMLYRV 497

RESULT 25

ABP39975

ID ABP39975 standard; protein; 1211 AA.

XX ABP39975;

XX 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4820.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.

XX Staphylococcus epidermidis.



XX PN US6380370-B1.  
 XX PD 30-APR-2002.  
 XX PF 13-AUG-1998; 98US-00134001.  
 XX PR 14-AUG-1997; 97US-0055779P.  
 XX PR 08-NOV-1997; 97US-0064964P.  
 XX PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX PI Doucette-Stamm LA, Bush D;  
 XX DR WPI; 2002-381255/41.  
 XX DR N-PSDB; ABN92520.  
 XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermis  
 XX PT polypeptide, useful for diagnosing and treating bacterial infections.  
 XX PS Disclosure; SEQ ID NO 4820; 267pp; English.  
 XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life cycle  
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site  
 XX XX Sequence 1211 AA;  
 Query Match 6.9%; Score 104; DB 5; Length 1211;  
 Best Local Similarity 18.0%; Pred. No. 17;  
 Matches 60; Conservative 60; Mismatches 131; Indels 82; Gaps 9;  
 QY 7 EOTVEVVK-----SAITADGALDYNNKLYDQVVPWKTFTETIKELSRFKQVYEQEASV 60  
 Db 185 EESAGVLYKKRAESIOQLDHTENLNARVEDILYDLGRVEPLKEBAIAKEYKQLSKE 244  
 QY 61 LVGDIKVLMSQDKYFEATQTVYEWK-----GVVTQLLSAYILLFDEYNE 106  
 Db 245 MEQSDVIVTVSDIDHYTEDNQRLDERLNHLKSQQAEGQQAQINQLQKY----- 295  
 QY 107 KASAKQDILIRLDDGVNKLNEAOKSLGSSQSFNNASGKLLALDSQLTND-----FS 160  
 Db 296 -KGKQQN-----DYDIEKLN---YELVKATENYEQLSGKLVLEERKQKQONQSETNARYE 345  
 QY 161 EKSSYFQSOVDRIKEAYAGAAAGVAGPFGLIISYSIAAGVIEGKLIPELNDRL-----K 216  
 Db 346 BELDNLESQIDSIRKEKAQNE-----KLLADLNKQKQQLNK 381  
 QY 217 AVQNFPTSLVTVVKQAKDIDAALKLATEIAAIGEIKTETETTRFYVDYDDLMLSLKLG 276  
 Db 382 EQVELESLLJISDEQHEKLEIKNSYYTLMSEQSVVNDIRFLEHTINENEAKSRDLS 441  
 QY 277 AAKGMINTCNEYQQ-----RHGKTKLLEV 300  
 Db 442 RLVEAFNQLKDIQTQNTQTKQKEYQSSKKSMEKV 474

RESULT 26  
 ABG31849  
 ID ABG31849 standard; protein; 2334 AA.  
 XX AC ABG31849;  
 XX DT 05-NOV-2002 (first entry)  
 XX DE Human kinase, MEK1.

XX KW Extracellular signal regulated kinase; hyperalgesia; surgery;  
 KW opiod withdrawal; pain sensitisation; analgesic; chronic pain; ERK;  
 XX MEK1; human; enzyme.  
 XX OS Homo sapiens.  
 XX PN WO200258687-A2.  
 XX PD 01-AUG-2002.  
 XX PF 25-JAN-2002; 2002WO-US002128.  
 XX PR 25-JAN-2001; 2001US-0264336P.  
 XX PA (TEXA) UNIV TEXAS SYSTEM.  
 XX PI Gutstein HB;  
 XX DR WPI; 2002-608414/65.  
 XX DR N-PSDB; ABK90804.  
 XX PT Reducing or reversing tolerance, physical dependence, hyperalgesia,  
 PT withdrawal symptoms, or pain sensitization in patients on analgesics for  
 PT chronic pain, comprises inhibition of the extracellular signal-regulated  
 XX kinase (ERK).  
 PS Disclosure; Page 156-161; 163pp; English.  
 XX CC The invention relates to a method of reducing or reversing tolerance,  
 CC reducing the risk of physical dependence or hyperalgesia, reducing the  
 CC symptoms of opiod withdrawal or inhibiting pain sensitisation in a  
 CC patient taking analgesics. The method comprises administering an  
 CC analgesic and an extracellular signal-regulated kinase (ERK) inhibitor  
 CC comprised in a formulation to reduce or reverse tolerance, risk of  
 CC physical dependence, hyperalgesia, symptoms of opiod withdrawal, or  
 CC inhibiting pain sensitisation in patients taking analgesics for chronic  
 CC pain or those undergoing surgery. The present sequence represents the  
 CC amino acid sequence of human MEK1 (not defined)  
 XX XX Sequence 2334 AA;  
 Query Match 6.9%; Score 104; DB 5; Length 2334;  
 Best Local Similarity 25.6%; Pred. No. 41;  
 Matches 60; Conservative 39; Mismatches 85; Indels 50; Gaps 13;  
 QY 78 EATQTV-YEWGCVTVQLLSAYILLFDEYNEKK-ASAQKDILIRILD---DGVNKLNEAOK 132  
 Db 1203 KATESYQDKDGNVTSVKDAYGTETYEYKKNNDVTYKMDTEGNVTDIAYDGLDAYSETDQ 1262  
 QY 133 SLLGSS-----QSFNN--ASGKLLALDSQLTNDIFS---EKSSY-FQSOVDRIKEAYAGAA 182  
 Db 1263 SGKSSAAVYDKYGNQIQSSKDLASNTILKDGSPKAGKSGWNLTASKDR-RKISVIADK 1321  
 QY 183 AGIVAGPPLGL-IISYSIAAGVIEGKLIPELNDRLKAVONFTSLVTVVKQAKDIDAALK 241  
 Db 1322 SGVLGSKALEVLSQSTSGTDHG-----YSSATQTV-----EL 1355  
 QY 242 KLATEIAAIGIKTETETTRFYVDYDDLMLSLKGAACKMINTC-NEYQQRHGK 294  
 Db 1356 EPNITYTUSGKIKITLAKSRAYFNID-----LRDKQKRQIWIHNEYSALAGK 1403

RESULT 27  
 ABP29330  
 ID ABP29330 standard; protein; 541 AA.  
 XX AC ABP29330;  
 XX DT 02-JUL-2002 (first entry)  
 XX XX Streptococcus polypeptide SEQ ID NO 7836.  
 XX DE

KW Streptococcus; GAS; CBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX  
OS Streptococcus pyogenes.  
XX

PN WO200234771-A2.

PD 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB004789.

XX 27-OCT-2000; 2000GB-00026333.

PR 24-NOV-2000; 2000GB-00028727.

PR 07-MAR-2001; 2001GB-00005640.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

XX Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C;  
PI Tettelin H;

XX WPI; 2002-352536/38.

DR N-PSDB; ABN69961.

XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.

XX Claim 1; Page 3916; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (i), nucleic acids encoding (i), ABN6044-ABN71526 and  
CC antibodies that bind (i) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (i) are used to determine whether a compound binds to  
CC biological sample. (i) is used to determine whether a compound binds to  
CC (i). A composition comprising (i) or a nucleic acid encoding (i), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (i) may be used to recombinantly produce (i) and may be  
CC used in gene therapy. Antibodies to (i) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins

XX Sequence 541 AA;

Query Match 6.9%; Score 103.5; DB 5; Length 541;  
Best Local Similarity 21.6%; Pred. No. 6.7;  
Matches 71; Conservative 51; Mismatches 142; Indels 65; Gaps 14;

QY 26 FYNKLDQVPMKTFD--ETIKELSRFK-QEYSQASVILVG-----DIKVLMDSDQK 75

DB 49 FYQYADK-----EKIDLSEARKRASELDISAYOKKAKELVAKELRREGKIVTRDDPTH 104

QY 76 FYEATQTVYEWCGVVTOLLISAVILLFDEYNEKASAKQDKILIRILLDDGWNKLINEAKXSLL 135

DB 105 QENADMISYN-LAMKTNALLELRNLIDLEMOELANGEHKLTKKFLDEGYRKETEFQAGLL 163

QY 136 SSSQSQFNAGSKLLALDSOLTNDP--SEKSSYFQSQVDRIKEAYAGAAAGIVAGPFGLI 193

DB 164 GLSVA-SQASVKSIA-DAVINANFKGAKWSNIDWRQDKLSRIISQSVQSAILKNGKLT 221

QY 194 I-----SYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQAKDI--DAA 239

DB 222 IARDIRREBFDVSASYAKRLAITEHARVQMEVGRLSMAENGAFMDILPEPRACDVCKDIA 281

QY 240 K-----LKLATEAALGEIKETETTRFYVDYD-----DLMLSL---- 273

DB 282 KHGPYHLDKWRIGENSPFPFPCRCALVGVDESQVATDR-QLDYNRRNNNIDLMAKTQSF 340  
QY 274 -----LKGAAKKWINTCNEYQQRHGKKTLL 297  
DB 341 IINKDVRVSARKKVGTRYDFWAQDNTKKI 369

RESULT 28

AAU38921

ID AAU38921 standard; protein; 660 AA.

XX AC AAU38921;

XX 11-SEP-2003 (revised)

DT 16-JAN-2002 (first entry)

XX C. trachomatis CT875 protein.

XX Chlamydia; sexually transmitted disease; PID; antibacterial;

KW pelvic inflammatory disease; antigen; trachoma; gynecological;

KW acute respiratory tract infection; atherosclerosis; male infertility;

XX coronary heart disease.

OS Chlamydia trachomatis; serovar E.

XX WO200181379-A2.

XX 01-NOV-2001.

XX 23-APR-2001; 2001WO-US013081.

XX 21-APR-2000; 2000US-0198853P.

XX 20-JUL-2000; 2000US-0219752P.

XX (CORI-) CORIYA CORP.

PI Bhatia A, Probst P, Stromberg EJ;

XX WPI; 2001-616771/71.

XX N-PSDB; AAS56995.

XX New polynucleotide for treating Chlamydia infections encodes a  
PT polynucleotides containing an immunogenic portion of a Chlamydia antigen.

XX Claim 3; Page 205-207; 208pp; English.

CC The invention relates to isolated polynucleotide encoding at least a  
CC partial Chlamydia protein which is an antigenic fragment, or the  
CC complements, fragments, homologues and variants, and antibodies raised  
CC against the antigenic proteins (or fragments). The nucleic acids,  
CC proteins and antibodies are used to diagnose and treat Chlamydia  
CC infections (e.g. a sexually transmitted disease, pelvic inflammatory  
CC disease (PID), acute respiratory tract infection, trachoma,  
CC atherosclerosis and coronary heart disease) in a patient, and in the  
CC treatment of male infertility. The compounds of the invention are also  
CC useful for detecting the presence of Chlamydia in a patient, and  
CC stimulating and/or expanding T cells specific for a Chlamydia protein.  
CC The present sequence represents a Chlamydia antigen. (Updated on 11-SEP-  
CC 2003 to standardise OS field)

XX Sequence 660 AA;

Query Match 6.9%; Score 103.5; DB 4; Length 660;  
Best Local Similarity 21.7%; Pred. No. 8.7;  
Matches 68; Conservative 48; Mismatches 119; Indels 79; Gaps 14;

QY 8 QTVEVKSATITAGALDFYNYKL-----DQV-----IPWKTFF 40

DB 155 EVANNIKKALBAQKDTIDKLNKLVTLQNNKSLTEVLKTTSDAQIPAINSOLEINKNSA 214

QY 41 DETIKELSRFKQEYSQASVILVGDIKVLMDSDQ--KYFEATQTVYEWCGVVTOLLISAYI 98

DB 215 DQIKDLER--QNTISYEA-VLTNAGEVIKASSEAGIKLQALQSIDV-AGDQSQ---AAV 267



isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 58016; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 1020 AA;

Query Match 6.9%; Score 103.5; DB 6; Length 1020;  
Best Local Similarity 19.8%; Pred. No. 15;  
Matches 65; Conservative 41; Mismatches 90; Indels 133; Gaps 12;  
3 GIFAEQTEVVK-----SAIETADG-----ALDFYKNKLDQVVPKTFDETIKLSR 49  
159 GIF-EEAGLVKQKQKKAQKLPETEDNLSRVQDIIEHEEQLTPLAAQSEAAKEFLR 217  
50 FQOEYSQ-BASVLVGDIKVLLMSQDKYFEATQTVWCVGVVQTLSSAYILLDFEYNEKK 108  
218 LKETLTQTDVSLMVAEIKTAKD-----WDNQK 245  
109 ASAQKDILIRILDGYNKLEA---QKSLGSSQGFNNASGKLLALDLSLTNDFSEKSSY 165  
246 AOLAKPNL-----ELGKLESIGQESILAKQKENAQADRLIEKNQVLLDSEKLQ 299  
166 FOSQVDRIRKEAYAGAAAGIVAGPFGLIITSYIAAGVTEGKLIPELNRLKAVNFFTSL 225  
300 TEGQKD-----VLQERTKHTQSSQBY 321  
226 SVTVKQAKDIDAAKLLATEAATGEIKETETTFYVDYDMLSLKLGAAK-----280  
322 QTSIAEAQK-----KVK-----HFELQESLMKAAAEKETEIQ 354  
281 -----MINTCNEYQ--QRHGKTKLLEVPD 302  
355 KAEANLIKQOELEKYQKSTKELLAEIRD 383

RESULT 31  
ABU23940  
ID ABU23940 standard; protein; 1163 AA.

XX  
AC ABU23940;  
XX

19-JUN-2003 (first entry)  
Protein encoded by Prokaryotic essential gene #9467.  
Antisense; prokaryotic essential gene; cell proliferation; drug design.  
Clostridium acetobutylicum.  
W0200277183-A2.  
03-OCT-2002.  
21-MAR-2002; 2002WO-US009107.  
21-MAR-2001; 2001US-00815242.  
06-SEP-2001; 2001US-00948993.  
25-OCT-2001; 2001US-0342923P.  
08-FEB-2002; 2002US-00072851.  
06-MAR-2002; 2002US-0362699P.  
(ELIT-) ELITRA PHARM INC.  
Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
WPI; 2003-029926/02.  
N-PSDB; ACA27810.  
New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.  
Claim 25; SEQ ID NO 51864; 1766pp; English.  
The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
Sequence 1163 AA;  
Query Match 6.9%; Score 103.5; DB 6; Length 1163;  
Best Local Similarity 21.7%; Pred. No. 18;  
Matches 81; Conservative 55; Mismatches 142; Indels 95; Gaps 14;  
1 MTGIFAEQTEVVKSAIETADG-----ALDFYKNKLDQVVPKTFDETIK-----ELS 48

QY

Db 112 ITGDEVLEEGAKSVNEKQBIIGLSLDDFTRTV--VLPQGFSEFLKJEGKERNMLE 169  
 Qy 49 RF--KQEYSQAS-----VLGGDIKVLMDSDQKYFEATQTVYEWCGVVT 91  
 Db 170 RLFNQYQDELSFKLARKIRKERKENVVLGELGYENINEDVLKERRELLKNNDFN 229  
 Qy 92 QLLSAYILLFDEYNEKKAQAOKDILI-----RILDPGVNKLAE---AQKSLIGSSQS--- 140  
 Db 230 EASKEYLKAEEYNEGKEVWGQLQIEBEKNRVRKDLMEKDEIDLKERARLGESSKVK 289  
 Qy 141 -----FNASGKLLALDSQLTN-----DFSEKSYFOSQVDRTRKEAYAGAA 182  
 Db 290 PYIDNYENTLQIDILKEQILSRENTMKAISLEKEDMEKLSIAKNKEK----- 339  
 Qy 183 AGIVAGPGLIISYIAAGVIEGKLIPELNDRLKAVQNFFTSLSV-----TVKQAKND 235  
 Db 340 ---ALPFRMKHHILDAIKEKOLLDNKLEKQLOKIEKLSLEASNKEBELIKQNKID 395  
 Qy 236 IDAAKLKLAETAAIGETETTRTFYVDYDMLSLKGAAKQMIN-----TCNEYQ 289  
 Db 396 IDSLTKIQNLKESKIDNLKVPBE---YKNKINEGIFELRNVDKELKHKNKGLGLDCDKFQ 451  
 Qy 290 ----ORHGKTYLL 298  
 Db 452 VDFEKAQSKKEML 464

RESULT 32  
 ADE63514  
 ID ADE63514 standard; protein; 1976 AA.  
 AC ADE63514;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Rat Protein Q9JLT0, SEQ ID NO 9458.  
 XX  
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
 XX  
 OS Rattus norvegicus.  
 XX  
 FN WO2003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 XX (GEO ) GEN HOSPITAL CORP.  
 PA (FARH ) BAYER AG.  
 XX  
 PI Woolf C, D'urso D, Befort K, Costigan M;  
 XX  
 DR WPI; 2003-268312/26.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 PS Claim 1; Page; 1017pp; English.  
 XX

CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence

CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1976 AA;

Query Match 6.9%; Score 103.5; DB 7; Length 1976;  
 Best Local Similarity 21.9%; Pred. No. 37;  
 Matches 61; Conservative 43; Mismatches 96; Indels 79; Gaps 10;  
 Qy 10 VEVVKSATETADGALDFVNYLDQV-----IPWKTFTDETIKEL-----SRFKQEYSQE 57  
 Db 1222 LEKNKQGLETDNKEACEVQLQVQKAESEHKRKKLDAQVQLHAKVSBGDRRLVELAEK 1281  
 Qy 58 ASVL---VGDIKVLLMDSQDKYFEATQTVYEWCGVTVQLLSAYILLFDEYNEKKAQAOKD 114  
 Db 1282 ANKLQNELDNVSTLLEAEKKGKMFAK---DAAGLESQLODTQELLOBETRQKLNLSR- 1337  
 Qy 115 ILIRILDGYNKLEAKQKSLGSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSQVDRIR 174  
 Db 1338 --IFQLSEKNSLQEQEE---EBEARNKLEKQVLAQSLADT-----KKKVD-- 1382  
 Qy 175 KEAYAGAAAGIVAGPFGGLIISYIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTVKQANK 234  
 Db 1383 -----DLGTIEG-----LEBAKKKLLK 1399  
 Qy 235 IDAAKLKLAETAAIGETETTRTFYVDYDMLSL 273  
 Db 1400 DVEALSQRLEKVLAYD--KLETKNRLQQLDLDTVDL 1436

RESULT 33  
 ADE63518  
 ID ADE63518 standard; protein; 1976 AA.  
 XX  
 AC ADE63518;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Rat Protein Q9JLT0, SEQ ID NO 9462.  
 XX  
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
 XX  
 OS Rattus norvegicus.  
 XX  
 FN WO2003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 XX (GEO ) GEN HOSPITAL CORP.  
 PA

PA (FARB ) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX New composition comprising two or more isolated polypeptides, useful for

XX preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat

XX or human polynucleotides or a polynucleotide which represents a fragment,

XX derivative or allelic variation of the nucleic acid sequence. Also

XX claimed are a vector comprising the novel polynucleotide, a host cell

XX comprising the vector, a method for identifying a nucleotide sequence

XX which is differentially regulated in an animal subjected to pain and a

XX kit to perform the method, an array, a method for identifying an agent

XX that increases or decreases the expression of the polynucleotide sequence

XX that is differentially expressed in neuronal tissue of a first animal

XX subjected to pain, a method for identifying a compound which regulates

XX the expression of a polynucleotide sequence which is differentially

XX expressed in an animal subjected to pain, a method for identifying a

XX compound that regulates the activity of one or more of the

XX polynucleotides, a method for producing a pharmaceutical composition, a

XX method for identifying a compound or small molecule that regulates the

XX activity in an animal of one or more of the polypeptides given in the

XX specification, a method for identifying a compound useful in treating

XX pain and a pharmaceutical composition comprising the one or more

XX polypeptides or their antibodies. The polynucleotide or the compound that

XX modulates its activity is useful for preparing a medicament for treating

XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene

XX therapy). The sequence presented is a rat protein (shown in Table 2 of

XX the specification) which is differentially expressed during pain. Note:

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic form directly from WIPO at

XX ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1976 AA;

Query Match 6.9%; Score 103.5; DB 7; Length 1976;

Best Local Similarity 21.3%; Pred. No. 37;

Matches 61; Conservative 43; Mismatches 96; Indels 79; Gaps 10;

QY 10 VEVVKSATFADGALFYNYKLDV-----IPKTFDETIKEL-----SRFKQYISOE 57

DB 1222 LEKNQGLTDTNDELACEVKVLOQVKAESHKFKKLDVQVQLHAKVSGDRLRVLEAK 1281

QY 58 ASVL---VGDIKVLMDSDQKYEATQTVYEWGCVVTQLLSAYILLDFEYNEKKAQAQKD 114

DB 1282 ANKLQELNDVSTLLEAEKKGKFKAK---DAAGLESQDQDTQLLEQETRKLNLSR- 1337

QY 115 ILIRILDGYNKLNKAEQKSLGSSQPNNAAGKLLALDSQLTNDPFEKSYFSQVDRLR 174

DB 1338 --IRQLEEKNSLQEQEE---BEEAKNLEKQVLAQSLADT-----KKYVDD-- 1382

QY 175 KEATAGAAAGVFPGLIISYSTAAGVIRGKLIPELNDRLKAVQNF\*SLSVTVKQANK 234

DB 1383 -----DLGTIEG-----LEBAKKKLLK 1399

QY 235 DIDAAKLKLATEPIAAIGETITETTRFYVDYDLMLSL 273

DB 1400 DVEALSQRLBEKVLAYD--KLEKTKNRLQQLDLDLTVDL 1436

RESULT 34

AAU14603

ID AAU14603 standard; protein; 5373 AA.

XX AAU14603;

XX 24-OCT-2001 (first entry)

XX Novel bone marrow polypeptide #2.

XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;

XX haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;

XX wound healing; nutritional supplement; immune disorder;

XX severe combined immunodeficiency; SCID.

XX Homo sapiens.

XX WO200157187-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US003782.

XX 03-FEB-2000; 2000US-00496914.

XX 20-JUN-2000; 2000US-00598075.

XX 19-JUL-2000; 2000US-00620325.

XX 30-NOV-2000; 2000US-0250683P.

XX (HYSE-) HYSEQ INC.

XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;

XX Ren F, Drmanac RT;

XX WPI; 2001-488875/53.

XX N-PSDB; AAS22908.

XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and

XX gene therapy.

XX Claim 10; Page 198-208; 392pp; English.

XX AAU14602-AAU14794 represent novel bone marrow polypeptides of the

XX invention. The proteins and corresponding coding sequences may be used in

XX the prevention, diagnosis and treatment of diseases associated with

XX inappropriate bone marrow polypeptide expression. For example, to treat

XX disorders associated with decreased expression by rectifying mutations or

XX deletions in a patient's genome that affect the activity of the

XX polypeptides by expressing inactive proteins or to supplement the

XX patient's own production of the polypeptides. Additionally, the nucleic

XX acids may be used to produce the polypeptides, by inserting the nucleic

XX acids into a host cell and culturing the cell to express the protein. The

XX nucleic acid and its complementary sequences may also be used as DNA

XX probes in diagnostic assays to detect and quantitate the presence of

XX similar nucleic acid sequences in samples, and therefore which patients

XX may be in need of restorative therapy. The proteins may also be used as

XX antigens in the production of antibodies against bone marrow proteins and

XX in assays to identify modulators of their expression and activity. The

XX anti-bone marrow protein antibodies and antagonists may also be used to

XX down regulate expression and activity. The antibodies may also be used as

XX diagnostic agents for detecting the presence of the protein in samples

XX (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins may be

XX used to regulate haematopoiesis activity, and consequently in the

XX treatment of myeloid or lymph cell disorders; in tissue regeneration,

XX such as wound healing; as a nutritional supplement; and in treatment of

XX immune disorders such as severe combined immunodeficiency (SCID)

XX SQ Sequence 5373 AA;

Query Match 6.9%; Score 103.5; DB 4; Length 5373;

Best Local Similarity 23.0%; Pred. No. 1.4e+02;

Matches 56; Conservative 49; Mismatches 106; Indels 33; Gaps 10;

QY 29 KYLDQVTPKTFDETIKELSRFKQYISOEASVLVDGDKVILMDSDQKYEATQTVYEWCG 88

DB 1369 KYISDAL--RRLEEEKVVEEKQEHVEKVKELGWSVTLARNTQK---ATSETKEST 1423

QY 89 VVTQLLSAYILLDFEYNEK---ASAQDKILIRILDGYNKLNKAEQKSLGSSQFN--- 142

DB 1424 DIEKAILQEQVLSBELTTKQVSEAIKTSQIFLAKHG-HKLSEKKEKQI--SEQLNALN 1480

QY 143 -----NASKLLALDSQLTNDSEKSSYFQSQVDRIKE-----AYAGAAAGIVAGP 189  
 Db 1481 KAYHDLCDGSANQLOQLAHQTEQKT--LQKQNTCHQOLEDLCSWVGQAEALAGH 1538  
 QY 190 FGLIISYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDAALKLA-TEIA 248  
 Db 1539 QGRTTQODLSA---LQKNQSDLKLDLQDDIQNRATSFATVVKDIEGFMEENQTKLSPRELT 1595  
 QY 249 AIGE 252  
 Db 1596 ALRE 1599

RESULT 35  
 ADJ68935  
 ID ADJ68935 standard; protein; 5373 AA.  
 AC ADJ68935;  
 XX  
 DT 06-MAY-2004 (first entry)  
 DE Human heat mitochondrial protein as a therapeutic target SeqID741.  
 KW mitochondrial; human; screening assay; diabetes mellitus;  
 KW Huntington's disease; osteoarthritis;  
 KW Leber's hereditary optic neuropathy; LHON;  
 KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
 KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
 KW neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;  
 KW osteopathic; ophthalmological; cytostatic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003087768-A2.  
 XX  
 PD 23-OCT-2003.  
 XX  
 PF 04-APR-2003; 2003WO-US010870.  
 XX  
 PR 12-APR-2002; 2002US-0372843P.  
 PR 17-JUN-2002; 2002US-0389987P.  
 PR 20-SEP-2002; 2002US-0412418P.  
 XX  
 XX (MITO-) MITOKOR.  
 PA (BUCK-) BUCK INST AGE RES.  
 XX  
 PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;  
 PI Warnock DE;  
 XX  
 DR WPI; 2003-845369/78.  
 XX  
 PT Identifying a mitochondrial target for drug screening assays and for  
 PT treating diseases associated with altered mitochondrial function,  
 PT comprises detecting a modified polypeptide in a sample and correlating  
 PT with the disease.  
 XX  
 PS Claim 1; SEQ ID NO 741; 180pp; English.  
 XX  
 CC This invention relates to novel mitochondrial targets that can be used  
 CC for therapeutic intervention in treating a disease associated with  
 CC altered mitochondrial function. Specifically, it refers to a method for  
 CC identifying proteins of the human heart mitochondrial proteome that are  
 CC useful for drug screening assays, as well as therapeutic targets. The  
 CC present invention describes a method for identifying such proteins that  
 CC can be used in the treatment of various diseases associated with altered  
 CC mitochondrial function including diabetes mellitus, Huntington's disease,  
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
 CC compositions have neuroprotective, nontropic, antidiabetic,  
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
 CC cytostatic activities. This polypeptide sequence is a human heart  
 CC mitochondrial protein of the invention.

XX SQ Sequence 5373 AA;  
 Query Match 6.9%; Score 103.5; DB 7; Length 5373;  
 Best Local Similarity 23.0%; Pred. No. 1.4e+02;  
 Matches 56; Conservative 49; Mismatches 106; Indels 33; Gaps 10;  
 QY 29 KYLDQVPMKTFDETIKELSRFQOEYASVLDGDIKVLMDSDQKYFEATQTVYEWCG 88  
 Db 1369 KYISDAL--RRLEBEKVVEBEKQEHVEKVELLIGWSTLARNTQK--ATSETKEST 1423  
 QY 89 VVTQLLSAYILLDFEYNEKK--ASAOKDILIRILDGWNKLNKAEKSLGSSQSFN--- 142  
 Db 1424 DIEKAILEQQVLSBELTTKQEVSEAIKTSQFLAKHG-HKLSEKEKQI--SEQLNALN 1480  
 QY 143 -----NASKLLALDSQLTNDSEKSSYFQSQVDRIKE-----AYAGAAAGIVAGP 189  
 Db 1481 KAYHDLCDGSANQLOQLAHQTEQKT--LQKQNTCHQOLEDLCSWVGQAEALAGH 1538  
 QY 190 FGLIISYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDAALKLA-TEIA 248  
 Db 1539 QGRTTQODLSA---LQKNQSDLKLDLQDDIQNRATSFATVVKDIEGFMEENQTKLSPRELT 1595  
 QY 249 AIGE 252  
 Db 1596 ALRE 1599

RESULT 36  
 ADH85709  
 ID ADH85709 standard; protein; 284 AA.  
 AC ADH85709;  
 XX  
 DT 22-APR-2004 (first entry)  
 DE Enterococcus faecalis polypeptide #189.  
 KW Enterococcus faecalis infection; transcription regulatory element;  
 KW antibacterial.  
 OS Enterococcus faecalis.  
 XX  
 PN US6617156-B1.  
 XX  
 PD 09-SEP-2003.  
 XX  
 PF 13-AUG-1998; 98US-00134000.  
 XX  
 PR 15-AUG-1997; 97US-0055778P.  
 XX  
 PA (DOUC/) DOUCETTE-STAMM L A.  
 PA (BUSH/) BUSH D.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX  
 DR WPI; 2003-895394/82.  
 DR N-PSDB; ADH82304.  
 XX  
 PT New nucleic acid comprising a sequence encoding an Enterococcus faecalis  
 PT polypeptide, useful for preparing a composition for diagnosing or  
 PT treating E. faecalis infection.  
 PS Disclosure; SEQ ID NO 3594; 193pp; English.  
 XX  
 CC The invention relates to Enterococcus faecalis polynucleotides and  
 CC polypeptides. The invention also relates to a recombinant expression  
 CC vector comprising a polynucleotide operably linked to a transcription  
 CC regulatory element, a cell comprising a recombinant vector, a method for  
 CC producing an E. faecalis polypeptide, an isolated nucleic acid comprising  
 CC a sequence not given in the specification, a recombinant vector  
 CC comprising the nucleic acid and a cell comprising the recombinant vector.  
 CC The polynucleotides can be used to detect the presence of E. faecalis in



CC a sample. The sequences are useful for preparing a composition for  
CC diagnosing or treating Enterococcus faecalis infection. This sequence  
CC represents an E. faecalis polypeptide of the invention.

XX SQ Sequence 284 AA;  
Query Match 6.8%; Score 103; DB 7; Length 284;  
Best Local Similarity 21.0%; Pred. No. 3.2; Indels 100; Gaps 13;  
Matches 67; Conservative 46; Mismatches 106; Indels 100; Gaps 13;  
QY 16 ALETADGALDFNKYLDQVWPKTDETIKEL-----SRFKQEQSASVLDGIKVL--- 68  
DB 6 AVKVGEGANFLEKNEK-----VDYTIKNFDDATGLYKALENGEADAIYDDVPVLGYA 59  
QY 69 -----LMSQD-----KYFEATQTVYVCGVVTQLLSAYILLFD 102  
DB 60 VKNQGKQLQVGDKEGTSSGYFAVKKQNPFLIKKFNAGLKNLKGNGTYDKILNNYLATGD 119  
QY 103 EYNEKKAQAQKDLIRILDDGVNKLNEAKSLGSSQS-----FNNAAGKLLALDSQLT 156  
DB 120 ETNTQDAGEQ-----MKITPKKEKYVIASDSTFAPFFQNAQGYVGDVLDV 168  
QY 157 NDFSEKSSV-----FQSQVDRIRKEAYAGAAAGIVAGPFLIISYSIAAGVIEGKLI 208  
DB 169 KEAAELQGFTEVEKFIGFSSAQAQVE-----SQADGMVAG-----M 205  
QY 209 PELNRLKA-----VQNFSTLSVTVKQAN---KDIDAAK-----LKLATEIA-AIGEIKT 255  
DB 206 TITDDRKAFAFDSVPYFDSGIQIAVKGNDKITSYDLKGGKVGKIGTESADFLERKKNK 265  
QY 256 ETETRFYVDYDMLSL 274  
DB 266 KYDYSIKYLDITDLYSAL 284

RESULT 37  
AAG82283  
ID AAG82283 standard; protein; 885 AA.  
XX AC AAG82283;  
XX DT 03-SEP-2001 (first entry)  
XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:1660.  
XX KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;  
XX OS endocarditis.  
XX PN Staphylococcus epidermidis.  
XX PD WO200134809-A2.  
XX PF 17-MAY-2001.  
XX PR 09-NOV-2000; 2000WO-US030782.  
XX PR 09-NOV-1999; 99US-0164258P.  
XX PA (GLAX ) GLAXO GROUP LTD.  
XX PI Kimmerly WJ;  
XX DR WPI; 2001-316495/33.  
XX DR N-PSDB; AAH53133.  
XX KW Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
XX PT useful for vaccinating against infections, e.g. endocarditis.  
XX PS Claim 18; Page 462; 2188pp; English.  
XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
XX CC (II), given in AAG81454 to AAG81120, from Staphylococcus epidermidis (I)  
XX CC and (II) can have antibacterial activity and therefore can be used in

CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
XX for SEQ ID NO:4455 to 4464

XX SQ Sequence 885 AA;  
Query Match 6.8%; Score 103; DB 4; Length 885;  
Best Local Similarity 18.2%; Pred. No. 14;  
Matches 53; Conservative 52; Mismatches 111; Indels 76; Gaps 8;  
QY 42 ETIKELSRFKQEQSASVLDGIKVLMSQDKYFEATQTVYVWC----- 87  
DB 2 EPLKEEAAIAKQKLSKEMEQQSDVIVTVSDIDHYTENDQRLDERNLNHLKSOQAEXEQQ 61  
QY 88 GVTQTLSAYILLFDEYNEKKAQAQKDLIRILDDGVNKLNEAKSLGSSQSFNNASGK 147  
DB 62 AQINQLLOKY-----KGRQON-----DYDIEKLN---YELVKATENYEQLSGK 102  
QY 148 LLALDSQLTND-----FSEKSSYFQSQVDRIKREAYAGAAAGIVAGPFLIISYSIAAG 201  
DB 103 LNVLEERKKQSETNARYEEELNLESQIDSINKKAQNE----- 142  
QY 202 VIEGKLIPELNRL-----KAVQNFSTLSVTVKQANKDIDAANKLATEIAAIGEIKTET 257  
DB 143 -----KLIAELKNKQKQLNKEVQELSLIYISDQHDKELEBEIKNSYYTLMSEQSDVNNDI 198  
QY 258 ETETRFYVDYDMLSLKGAQAKMINTCNEYQQ-----RHGKKTLLLEV 300  
DB 199 RFLHTINEAKKSRDLSDRLVEAFNQLKDIOQNIQTOTQKEYOSSKSKMEKV 250

RESULT 38  
ADB67135  
ID ADB67135 standard; protein; 961 AA.  
XX AC ADB67135;  
XX DT 04-DEC-2003 (first entry)  
XX DE General vesicular transport factor p115 SEQ ID NO:155.  
XX KW staged assembly; nanostructure; peptide nucleic acid; PNA;  
XX KW structural reinforcement; aerogel; paper; plastic; cement;  
XX KW tensile strength; identification marker; anti-counterfeiting marker;  
XX KW enzyme support; catalyst support; assembly scaffold; nanowire;  
XX KW nanocircuit; molecular sieve; molecular filter; biosensor.  
XX OS Bos taurus.  
XX PN WO2003072829-A1.  
XX PD 04-SEP-2003.  
XX PF 21-FEB-2003; 2003WO-US005390.  
XX PR 21-FEB-2002; 2002US-00080608.  
XX PA (NANO-) NANOFRAMES INC.  
XX CC



PI Hyman PL, Goldberg EB;  
XX WPI; 2003-721788/68.  
XX Staged assembly of nanostructures, useful e.g. in biosensors or as  
PT catalyst supports, using assembly units derived from peptide nucleic  
PT acids.  
XX  
XX Disclosure; Page 58; 118pp; English.  
XX  
CC The present invention describes a method (M1) for the staged assembly of  
CC a nanostructure using peptide nucleic acids (PNAs). M1 comprises: (a)  
CC contacting a nanostructure intermediate (NSI) having at least one unbound  
CC joining element (JE) with an assembly unit (AU) that comprises several  
CC different JE where: (i) none of these JE can interact with itself or  
CC other JE; and (ii) only one JE in AU and a single unbound JE in NSI are  
CC complementary, so that AU becomes non-covalently linked to NSI to produce  
CC a new NSI for use in subsequent cycles; (b) removing unbound AU; and (c)  
CC cyclic repetition of (a) and (b) to form a nanostructure. The new feature  
CC is that the complementary JE in at least one cycle are PNAs. Also  
CC described are nanostructures formed from many AU, comprising different  
CC JE, where at least one AU includes PNA. M1 is useful for producing  
CC nanostructures with a very wide range of potential applications, e.g.  
CC structural reinforcements (for aerogels, paper, plastics or cement,  
CC particularly as long fibres to improve tensile strength); identification  
CC (anti-counterfeiting) markers; enzyme or catalyst supports; assembly  
CC scaffolds; for construction of nanowires or nanocircuits; size markers  
CC for electron microscopy; molecular sieves and filters; substrates for  
CC optical and other surface coatings; scaffolds for solubilising enzymes or  
CC for trapping, protecting and delivering specific molecules; in high-  
CC density computer memories; as artificial zeolite for absorbing ions from  
CC water and for construction of new materials, including use in biosensors.  
CC PNAs are more homogeneous than inorganic nanoparticles generally used to  
CC form nanostructures, so will produce structures with predictable geometry  
CC and stoichiometry. The present sequence represents a protein containing  
CC coiled coil dimerisation sequences that can be used for structural  
CC elements of assembly units, given in the exemplification of the present  
XX invention.  
XX  
SQ Sequence 961 AA;

Query Match 6.8%; Score 103; DB 7; Length 961;  
Best Local Similarity 22.0%; Pred. No. 16;  
Matches 70; Conservative 53; Mismatches 129; Indels 66; Gaps 14;  
QY 11 EVKSAIETADGALDFYKNYL-----DQVTPKTFDE-----TIKELSRFK 51  
DB 639 EVKKTLEQHSIVTHYKNMIREQDLQLEELKQKISTLKQNEQLQTAVTQVQSQIQQHK 698  
QY 52 QEYSQASVLVGDIKVLL-MDSQDK--YFEATQTVYEWCGVVTQLLSAYILLFDEYNEKK 108  
DB 699 DQYNL-----LKVQLGKDSHQGPYTDGAQ----MNGVQPEISR--LREEIEELK 743  
QY 109 AS-----AQKDLIRLLDDGVNK--LNEAQKLLGSSQSFNNASKLLALSQLTND 158  
DB 744 SNRELQSLQAEKDSLIENLKSQSLSPTNEQSSATAGDSEQIABLKQELATLKSQ--NS 802  
QY 159 FSEKSSYFQSOVDRTFRKAYAGAAAGIVAGPGLIISYIAAGVTEGKL-----IPELN 212  
DB 803 QSVETIKLQTEKQELQKTEAFKAPVPGSEVTATKTTD--VEGRSLALQETKELK 860  
QY 213 DRLKAVQNFFTSLSVTVKQANKDIDAALKLKLATEIAAIGEIKTETETTRFYVDVDDLMS 272  
DB 861 NEIKALSERTAIKQLDSSNSTI-----AILQNEKNKLEVDITDSKQEQDILLV- 910  
QY 273 LLKGAKKKMNCTENYQQ 290  
DB 911 LLADQDQKIFSLKKNLKE 928  
RESULT 39  
ADL99406  
ID ADL99406 standard; protein; 961 AA.

XX  
AC ADL99406;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE Nanostructure assembly protein #57.  
XX  
KW peptide nucleic acid; PNA; nanostructure.  
XX  
OS Synthetic.  
XX  
FN US2003215903-A1.  
XX  
PD 20-NOV-2003.  
XX  
PF 21-FEB-2003; 2003US-00370685.  
XX  
PR 21-FEB-2002; 2002US-00080608.  
XX  
PA (HYMA)/ HYMAN P L.  
PA (GOLD)/ GOLDBERG E B.  
XX  
PI Hyman PL, Goldberg EB;  
XX  
DR WPI; 2004-021840/02.  
XX  
XX Staged assembly of a nanostructure containing peptide nucleic acid  
PT assembly units comprising a nanostructure intermediate with an  
PT assembly unit comprising different joining elements, and removing unbound  
PT assembly units.  
XX  
PS Disclosure; Page 43-44; 73pp; English.  
XX  
CC The invention relates to staged assembly of a nanostructure comprising:  
CC (a) contacting a nanostructure intermediate comprising at least one  
CC unbound joining element with an assembly unit comprising different  
CC joining elements; (b) removing unbound assembly units; and (c) repeating  
CC steps (a)-(b) for a sufficient number of cycles to form a nanostructure,  
CC where the assembly unit in at least one cycle comprises a peptide nucleic  
CC acid. A single joining element of the different joining elements and a  
CC single unbound joining element of the nanostructure intermediate are  
CC complementary joining elements, where the assembly unit is non-covalently  
CC bound to the nanostructure intermediate to form a new nanostructure  
CC intermediate for use in subsequent cycles. The method for staged assembly  
CC of a nanostructure further comprises: (i) capping the nanostructure with  
CC at least one capping unit; and (ii) post-assembly conversion of specific  
CC non-covalent interactions of complementary joining elements to covalent  
CC linkages, where the linkages are stabilised. The nanostructure  
CC intermediate comprises a surface bound initiator assembly unit. A first  
CC assembly unit used in at least one cycle comprises at least one  
CC structural element covalently linked to a first joining element  
CC comprising a peptide nucleic acid, or a first structural element bound to  
CC a second structural element to form a stable complex. The structural  
CC element is covalently linked to the first joining element and to a second  
CC joining element, comprising a peptide nucleic acid. The assembly unit  
CC further comprises a functional element, comprising photoactive molecule,  
CC photonic nanoparticle, inorganic ion, carbon nanotube, lectin or  
CC chemiluminescent molecule. The assembly unit comprises sub-assembly units  
CC that bind to each other to form a complex. The present sequence  
CC represents a protein containing a coiled coil dimerisation sequence that  
CC can be used as a structural element in the method of the invention.  
XX  
SQ Sequence 961 AA;

Query Match 6.8%; Score 103; DB 8; Length 961;  
Best Local Similarity 22.0%; Pred. No. 16;  
Matches 70; Conservative 53; Mismatches 129; Indels 66; Gaps 14;  
QY 11 EVKSAIETADGALDFYKNYL-----DQVTPKTFDE-----TIKELSRFK 51  
DB 639 EVKKTLEQHSIVTHYKNMIREQDLQLEELKQKISTLKQNEQLQTAVTQVQSQIQQHK 698  
QY 52 QEYSQASVLVGDIKVLL-MDSQDK--YFEATQTVYEWCGVVTQLLSAYILLFDEYNEKK 108

Db 699 DQYNL-----LKVQKDSQHOGPTDGAQ-----MNGVQPEISR---LREIEELK 743

QY 109 AS-----AQKOLIRLDDGVNK--LNEAQKSLGSSQSFNNASGKLALDSQLTND 158

Db 744 SNRELQSLAEKDSLIENKSSQLSPGTNEQSSATAGDSQIAELKQELATLKSQ--NS 802

QY 159 FSEKSSYFQSDVRIRKRYAGAAAGIVAGFPGLIISYIAAGVIEGKL-----IPELN 212

Db 803 QSVETIKLQTEKQELLQKTEAFKAPVPGSESVIATKTTD--VEGRSLALLOETKELK 860

QY 213 DLKKAQVNFSTLSVTVKQANKDIDAAKLKATEAAICEIKTETETTFRFVYDVLMLIS 272

Db 861 NEIKALSBEERTAIKQLODSSNTI-----AILQNEKKNLEVDITDSKKEQDILLV- 910

QY 273 LLKGAAKKWMINTCNEYQQ 290

Db 911 LLADQDQKIFSLKNKLKE 928

RESULT 40

ADJ69829

ID ADJ69829 standard; protein; 1583 AA.

XX AC ADJ69829;

XX DT 06-MAY-2004 (first entry)

XX DE Human heat mitochondrial protein as a therapeutic target-SeqID1635.

XX KW mitochondrial; human; screening assay; diabetes mellitus;

XX KW Huntington's disease; osteoarthritis;

XX KW Leber's hereditary optic neuropathy; LHON;

XX KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;

XX KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;

XX KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;

XX KW osteopathic; ophthalmological; cytostatic.

XX OS Homo sapiens.

XX PN WO2003087768-A2.

XX PD 23-OCT-2003.

XX PF 04-APR-2003; 2003WO-US010870.

XX PR 12-APR-2002; 2002US-0372843P.

XX PR 17-JUN-2002; 2002US-0389987P.

XX PR 20-SEP-2002; 2002US-0412418P.

XX PA (MITO-) MITOKOR

XX PA (BUCK-) BUCK INST AGE RES.

XX PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;

XX PI Warnock DE;

XX XX WPI; 2003-845369/78.

XX XX Identifying a mitochondrial target for drug screening assays and for

XX PT treating diseases associated with altered mitochondrial function,

XX PT comprises detecting a modified polypeptide in a sample and correlating

XX PT with the disease.

XX XX Claim 1; SEQ ID NO 1635; 180pp; English.

XX PS This invention relates to novel mitochondrial targets that can be used

XX CC for therapeutic intervention in treating a disease associated with

XX CC altered mitochondrial function. Specifically, it refers to a method for

XX CC identifying proteins of the human heart mitochondrial proteome that are

XX CC useful for drug screening assays, as well as therapeutic targets. The

XX CC present invention describes a method for identifying such proteins that

XX CC can be used in the treatment of various diseases associated with altered

XX CC mitochondrial function including diabetes mellitus, Huntington's disease,

CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial

CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy

CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these

CC compositions have neuroprotective, nootropic, antidiabetic,

CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and

CC cytostatic activities. This polypeptide sequence is a human heart

CC mitochondrial protein of the invention.

XX XX Sequence 1583 AA;

QY Query Match 6.8%; Score 103; DB 7; Length 1583;

Db Best Local Similarity 19.9%; Pred. No. 30;

QY Matches 73; Conservative 55; Mismatches 89; Indels 150; Gaps 16;

QY 7 EQTEVVVKSATADGALDFYKYLQVVPWKT--FDETIKELSRFKQEYSQASVVLVD 64

Db 531 EQTIQ-----YNSLEKQKVNELTGLEETLKE---KDQNDQKLEKLMVQ 571

QY 65 IKVLLMDSQDYFEATQTVYEMGCVTQLLSAVI--LLFDYNEKKASAKDILIRILD-- 121

Db 572 MKVL---SEDK-----EVLASAEVKSLYEENN--KLSEKKQLSRDLLEV 610

QY 122 -----DGVNKL--NEAOKSLGSSQSFNNASGKLIALD 152

Db 611 LSQKEDVILKEHITOLEKKQLMVEEQDNKLENEQVKLFVKTLQY---GFLKEMG 666

QY 153 SOLTNDSEKSSYFQSDVRIRKEAYAGAAAGIVAGFPGLIISYIAAGVIEGKL----- 207

Db 667 SEVSEDSKED-----VAVNLQAVGSLAKINEEK 697

QY 208 -----IPELNDRLKAVQNFTSLSVTVKQANKDIDAAKLKATEIAAIG----- 251

Db 698 NLAFQDEKVELEKEIKCLQESVWQCBELSLRDYEQEKVLLRKELEIQSEKALQ 757

QY 252 ----EIKTETETTFYVDYDMLSLKGAAKWMINTCNEYQQRHGK-----KT 296

Db 758 SDLEKMNANEKTR--LENQNLLIQV-----EEVSOICSKSEIHNEKEKCFIKEHENLKP 810

QY 297 LLEVPDI 303

Db 811 LLEQKEL 817

RESULT 41

ABB59344

ID ABB59344 standard; protein; 2056 AA.

XX AC ABB59344;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 4824.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL03447.



RESULT 43  
AAG46982  
ID AAG46982 standard; protein; 746 AA.  
XX AC AAG46982;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 59165.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140931P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145244P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.

|           |                               |  |     |
|-----------|-------------------------------|--|-----|
| Db        | 162                           | LERTRQANEALKAMDAERQQLRSANKKLRTDIEELRGLSLQPKENKIETLQOSLLDKQOI             | 221 |
| Qy        | 173                           | ---IRKEAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTV               | 229 |
| Db        | 222                           | LEDLKKQLQAVEERKQIA-VTELSAKHOKNLEGLAQVVDALSERDKAAET-ISSLQVLL              | 279 |
| Qy        | 230                           | KQANKDI-----DAANKLATEI--AAIGEIKTETE                                      | 258 |
| Db        | 280                           | AEKESKIAEMEAAATGEAARLRAAAETLKGELAHUKSENE                                 | 319 |
| RESULT 44 |                               |  |     |
| AA646981  |                               |  |     |
| ID        | AA646981                      | standard; protein; 788 AA.   |     |
| XX        | AA646981;                     |  |     |
| XX        |                               |  |     |
| DT        | 18-OCT-2000                   | (first entry)  |     |
| XX        |                               |  |     |
| DE        | Arabidopsis thaliana          | protein fragment SEQ ID NO: 59164.                                       |     |
| XX        |                               |  |     |
| KW        |                               | Protein identification; signal transduction pathway; metabolic pathway;  |     |
| KW        |                               | hybridisation assay; genetic mapping; gene expression control; promoter; |     |
| KW        |                               | termination sequence.  |     |
| XX        |                               |  |     |
| OS        | Arabidopsis thaliana.         |  |     |
| XX        |                               |  |     |
| PN        | EP1033405-A2.                 |  |     |
| XX        |                               |  |     |
| PD        | 06-SEP-2000.                  |  |     |
| XX        |                               |  |     |
| PF        | 25-FEB-2000; 2000EP-00301439. |  |     |
| XX        |                               |  |     |
| PR        | 25-FEB-1999;                  | 99US-0121825P.   |     |
| PR        | 03-MAR-1999;                  | 99US-0123180P.   |     |
| PR        | 03-MAR-1999;                  | 99US-0123548P.   |     |
| PR        | 23-MAR-1999;                  | 99US-0125788P.   |     |
| PR        | 23-MAR-1999;                  | 99US-0126264P.   |     |
| PR        | 23-MAR-1999;                  | 99US-0126785P.   |     |
| PR        | 01-APR-1999;                  | 99US-0127462P.   |     |
| PR        | 06-APR-1999;                  | 99US-0128234P.   |     |
| PR        | 08-APR-1999;                  | 99US-0128714P.   |     |
| PR        | 16-APR-1999;                  | 99US-0129845P.   |     |
| PR        | 19-APR-1999;                  | 99US-0130077P.   |     |
| PR        | 21-APR-1999;                  | 99US-0130449P.   |     |
| PR        | 23-APR-1999;                  | 99US-0130510P.   |     |
| PR        | 23-APR-1999;                  | 99US-0130891P.   |     |
| PR        | 23-APR-1999;                  | 99US-0131449P.   |     |
| PR        | 30-APR-1999;                  | 99US-0132048P.   |     |
| PR        | 30-APR-1999;                  | 99US-0132407P.   |     |
| PR        | 04-MAY-1999;                  | 99US-0132484P.   |     |
| PR        | 05-MAY-1999;                  | 99US-0132485P.   |     |
| PR        | 06-MAY-1999;                  | 99US-0132486P.   |     |
| PR        | 06-MAY-1999;                  | 99US-0132487P.   |     |
| PR        | 07-MAY-1999;                  | 99US-0132863P.   |     |
| PR        | 11-MAY-1999;                  | 99US-0134256P.   |     |
| PR        | 14-MAY-1999;                  | 99US-0134218P.   |     |
| PR        | 14-MAY-1999;                  | 99US-0134219P.   |     |
| PR        | 14-MAY-1999;                  | 99US-0134221P.   |     |
| PR        | 14-MAY-1999;                  | 99US-0134370P.   |     |
| PR        | 18-MAY-1999;                  | 99US-0134768P.   |     |
| PR        | 19-MAY-1999;                  | 99US-0134941P.   |     |
| PR        | 20-MAY-1999;                  | 99US-0135124P.   |     |
| PR        | 21-MAY-1999;                  | 99US-0135353P.   |     |
| PR        | 22-MAY-1999;                  | 99US-0135629P.   |     |
| PR        | 23-MAY-1999;                  | 99US-0136021P.   |     |
| PR        | 27-MAY-1999;                  | 99US-0136392P.   |     |
| PR        | 28-MAY-1999;                  | 99US-0136782P.   |     |
| PR        | 01-JUN-1999;                  | 99US-0137222P.   |     |
| PR        | 03-JUN-1999;                  | 99US-0137528P.   |     |
| PR        | 04-JUN-1999;                  | 99US-0137502P.   |     |
| PR        | 07-JUN-1999;                  | 99US-0137724P.   |     |
| PR        | 08-JUN-1999;                  | 99US-0138094P.   |     |

Query Match

Best Local Similarity 22.5%; Pred. No. 12;

Matches 63; Conservative 51; Mismatches 121; Indels 45; Gaps 12;

Qy

2

TGFAEQTEVVVKSAIETADGALDPYKYLQVLPWKTFDETIKELSRFKQYSGEASVL 61

Db

62

TRVAEQALEHLREAYSEADAKSQYSKFSQV--EQKLDQEIHERD---EKYADL----- 112

Qy

62

VGDIKVLVMSODKYFEATQTVYECGVVVTQLLSAYILLFDEYNE--KKASAAQKDILIRI 119

Db

113

--DAKFTLHKR-----AKQIQE-----IQEKODLDARFREVNETAERASSQHSMMQE 161

Qy

120

LDDGVNKLNEAKSLGSSQFNASGKLALDLSQLTWDFSEKSYF-----OSQVDR--- 172

|           |  |   |
|-----------|--|---|
| Db        | 162  | LERTQOANEALKAMDAEROQLRSANNKLRDITIELRGLSQPKENKIETLQOSSLDDKQI 221 |
| Qy        | 173  | ---IRKEAYAGAAAGIVAGPFGLLIYSIAAGVIEGKLIPELNDKRLKAVQNFSTLSVTV 229 |
| Db        | 222  | LEDLKQLQAVEERKQIA-VTELSAKHOKNLEGEAQVVDALSERDKAAET-ISSLQVLL 279  |
| Qy        | 230  | KQANKDI-----DAAKLKLAETI--AAIGIKTETE 258                         |
| Db        | 280  | AEKESKIAEMEAATGEAARLRAAAATLKGELAHKSENE 319                      |
| RESULT 44 |  |   |
| AAG46981  |  |   |
| ID        | AAG46981 standard; protein; 788 AA.                                      |   |
| XX        | AAC46981;  |   |
| XX        | 18-OCT-2000 (first entry)  |   |
| XX        | Arabidopsis thaliana protein fragment SEQ ID NO: 59164.                  |   |
| XX        | Protein identification; signal transduction pathway; metabolic pathway;  |   |
| XX        | hybridisation assay; genetic mapping; gene expression control; promoter; |   |
| XX        | termination sequence.  |   |
| XX        | Arabidopsis thaliana.  |   |
| XX        | EP1033405-A2.  |   |
| XX        | 06-SEP-2000.   |   |
| XX        | 25-FEB-2000; 2000EP-00301439.  |   |
| XX        | 25-FEB-1999; 99US-0121825P.  |   |
| XX        | 05-MAR-1999; 99US-0123180P.  |   |
| XX        | 09-MAR-1999; 99US-0123548P.  |   |
| XX        | 23-MAR-1999; 99US-0125788P.  |   |
| XX        | 25-MAR-1999; 99US-0126264P.  |   |
| XX        | 29-MAR-1999; 99US-0126785P.  |   |
| XX        | 01-APR-1999; 99US-0127462P.  |   |
| XX        | 08-APR-1999; 99US-0128234P.  |   |
| XX        | 08-APR-1999; 99US-0128714P.  |   |
| XX        | 16-APR-1999; 99US-0129845P.  |   |
| XX        | 19-APR-1999; 99US-0130077P.  |   |
| XX        | 21-APR-1999; 99US-0130449P.  |   |
| XX        | 23-APR-1999; 99US-0130510P.  |   |
| XX        | 28-APR-1999; 99US-0130891P.  |   |
| XX        | 30-APR-1999; 99US-0131449P.  |   |
| XX        | 30-APR-1999; 99US-0132048P.  |   |
| XX        | 04-MAY-1999; 99US-0132407P.  |   |
| XX        | 05-MAY-1999; 99US-0132484P.  |   |
| XX        | 06-MAY-1999; 99US-0132485P.  |   |
| XX        | 06-MAY-1999; 99US-0132486P.  |   |
| XX        | 07-MAY-1999; 99US-0132487P.  |   |
| XX        | 11-MAY-1999; 99US-0132863P.  |   |
| XX        | 14-MAY-1999; 99US-0134256P.  |   |
| XX        | 14-MAY-1999; 99US-0134218P.  |   |
| XX        | 14-MAY-1999; 99US-0134219P.  |   |
| XX        | 14-MAY-1999; 99US-0134221P.  |   |
| XX        | 14-MAY-1999; 99US-0134370P.  |   |
| XX        | 18-MAY-1999; 99US-0134768P.  |   |
| XX        | 19-MAY-1999; 99US-0134941P.  |   |
| XX        | 20-MAY-1999; 99US-0135124P.  |   |
| XX        | 21-MAY-1999; 99US-0135353P.  |   |
| XX        | 25-MAY-1999; 99US-0135629P.  |   |
| XX        | 27-MAY-1999; 99US-0136021P.  |   |
| XX        | 28-MAY-1999; 99US-0136782P.  |   |
| XX        | 01-JUN-1999; 99US-0137222P.  |   |
| XX        | 03-JUN-1999; 99US-0137528P.  |   |
| XX        | 04-JUN-1999; 99US-0137502P.  |   |
| XX        | 07-JUN-1999; 99US-0137724P.  |   |
| XX        | 08-JUN-1999; 99US-0138094P.  |   |

PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 21-JUN-1999; 99US-0139899P.  
PR 22-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 24-JUN-1999; 99US-0140823P.  
PR 28-JUN-1999; 99US-0140991P.  
PR 29-JUN-1999; 99US-0142803P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 01-JUL-1999; 99US-0142055P.  
PR 02-JUL-1999; 99US-0142390P.  
PR 06-JUL-1999; 99US-0142803P.  
PR 08-JUL-1999; 99US-0142920P.  
PR 09-JUL-1999; 99US-0142977P.  
PR 12-JUL-1999; 99US-0143542P.  
PR 13-JUL-1999; 99US-0144005P.  
PR 14-JUL-1999; 99US-0144085P.  
PR 15-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 20-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 21-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 22-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 23-JUL-1999; 99US-0145276P.  
PR 26-JUL-1999; 99US-0145313P.  
PR 27-JUL-1999; 99US-0145318P.  
PR 27-JUL-1999; 99US-0145319P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158322P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 18-OCT-1999; 99US-0159638P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.8%; Score 102.5; DB 3; Length 788;  
Best Local Similarity 22.5%; Pred No. 13;  
Matches 63; Conservative 51; Mismatches 121; Indels 45; Gaps 12;  
2 TGIFAEQTVVVKSAIETADGALDFYKYLDOVIPWTFDETIKELSRFKQESQASVL 61

db 104 TRVAAEQALEHLREAYSEADAKSQRYSKFSQV--EQKLDQBIKRD---EKYADL----- 154  
QY 62 VGDIKVLLMDSQDKYFEATQTVYVCGVVTQLLSAYILLFDEYNB--KKAQAQDKILIRI 119  
Db 155 --DAFTRLHR-----AKRIQE-----IQEKDLDARFREVNTAERASSQHSQMOE 203  
QY 120 LDDGVNKLNEAQKSLGSSQFNNAAGKLLALDSQLTNDFFSKSYF-----QSQVDR--- 172  
Db 204 LERTQQAQNEALKAMDAERQQLRSANNKLRDTIELRGLSLOPKENKIETLQQLSLDKQI 263  
QY 173 ---IRKEAYAGAAAGIVACPGFLIISYSIAAGVIEGKLIPELNDRLKAVQNFFTSLSYTV 229  
Db 264 LEDLKKQLQAVEERKQIA-VTELSAKHQKNGLEQAQVVDALSERDKAAET-ISSLQVLL 321  
QY 230 QKANKDI-----DAAKKLATEI--AAIGEIKTETE 258  
Db 322 AEKESKIAEMEAAATGEAARLARARAAATLKGELAHLSENE 361

RESULT 45

AAU14697  
ID AAU14697 standard; protein; 5447 AA.  
AC AAU14697;  
XX  
XX 24-OCT-2001 (first entry)  
XX Novel bone marrow polypeptide #96.  
XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;  
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;  
KW wound healing; nutritional supplement; immune disorder;  
KW severe combined immunodeficiency; SCID.

XX Homo sapiens.  
XX WO200157187-A2.  
XX 09-AUG-2001.  
XX 05-FEB-2001; 2001WO-US003782.  
XX 03-FEB-2000; 2000US-00496914.  
XX 20-JUN-2000; 2000US-00598075.  
XX 19-JUL-2000; 2000US-00620325.  
XX 30-NOV-2000; 2000US-0250683P.  
XX (HYSE-) HYSEQ INC.  
PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;  
PI Ren F, Drmanac RT;  
XX  
DR WPI: 2001-488875/53.  
DR N-PSDB; AA523002.  
XX  
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and gene therapy.

XX Claim 10; Page 124-127; 392pp; English.  
XX AAU14602-AAU14794 represent novel bone marrow polypeptides of the invention. The proteins and corresponding coding sequences may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate bone marrow polypeptide expression. For example, to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of the polypeptides by expressing inactive proteins or to supplement the patient's own production of the polypeptides. Additionally, the nucleic acids may be used to produce the polypeptides, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The nucleic acid and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of

CC similar nucleic acid sequences in samples, and therefore which patients may be in need of restorative therapy. The proteins may also be used as antigens in the production of antibodies against bone marrow proteins and in assays to identify modulators of their expression and activity. The anti-bone marrow protein antibodies and antagonists may also be used to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the protein in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins may be used to regulate haematopoiesis activity, and consequently in the treatment of myeloid or lymph cell disorders; in tissue regeneration, such as wound healing; as a nutritional supplement; and in treatment of immune disorders such as severe combined immunodeficiency (SCID)  
XX  
SQ Sequence 5447 AA;

Query Match 6.8%; Score 102.5; DB 4; Length 5447;  
Best Local Similarity 23.0%; Pred. No. 1.7e+02;  
Matches 56; Conservative 49; Mismatches 106; Indels 33; Gaps 10;  
QY 29 KYLDQVIPWKTFTETIKELSRFQESASVLGDIKVLMLDSQDKYFEATQTVYVCG 88  
Db 1421 KYISDAL--RRLEBEKVVBEKQEHVKEKLLGWSTLARNTQK---ATSETKEST 1475  
QY 89 VVTQLLSAYILLFDEYNEK---ASAQKDIILIRLDDGVNKLNEAQKSLGSSQSFN--- 142  
Db 1476 DIEKAILLEQQVLSBELTTKQEVSEAIKASQIFLAKHG-HKLSEKEKKQI--SEQLNALN 1532  
QY 143 -----NASGKLLALDSQLTNDFFSEKSYFSQSVDRIRKE-----AYAGAAAGIVAGP 189  
Db 1533 KAYHDLCDGSANQLQQLQSOLAHQTEKT--LQKQNTCHQQLDLCWSVGQAEALAGH 1590  
QY 190 FGLIISYSIAAGVIEGKLIPELNDRLKAVQNFFTSLSYTVKQANKDIDAQKLA-TEIA 248  
Db 1591 QGRTTQQLDSA---LQKNQSDLKDLQDDIQNRATSFATVWKDIEGFMEENQTKLSPRELT 1647  
QY 249 AIGE 252  
Db 1648 ALRE 1651

RESULT 46

AAE10039  
ID AAE10039 standard; protein; 590 AA.  
XX  
AC AAE10039;  
XX 11-SEP-2003 (revised)  
DT 29-NOV-2001 (first entry)  
XX  
XX N. meningitidis strain 2996 961c-741 fusion protein.  
XX Heterologous expression; Neisserial protein; 961c-741 fusion protein.  
XX Neisseria meningitidis; 2996.  
XX WO200164920-A2.  
XX 07-SEP-2001.  
XX 28-FEB-2001; 2001WO-IB000420.  
XX 28-FEB-2000; 2000GB-00004695.  
PR 13-NOV-2000; 2000GB-00027675.  
XX (CHIR-) CHIRON SPA.  
XX Arico'NB, Comanducci M, Galeotti C, Masignani V, Giuliani MM;  
PI Pizza M;  
XX WPI: 2001-557776/62.  
DR N-PSDB; AAD17054.  
XX Heterologous expression for the expression of two or more Neisserial









GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:07:04 ; Search time 41 Seconds  
(without alignments)

711.065 Million cell updates/sec

Title: US-09-993-292B-24

Perfect score: 1510

Sequence: 1 MTGIFAEQTEVVKSAIETA.....TCNEYQQRHGKTLLEVPDI 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR\_79.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 1472  | 97.5        | 305    | 2 AE0673 | haemolysin HlyE [i |
| 2          | 1381  | 91.5        | 305    | 2 C64864 | hemolysin E - Esch |
| 3          | 1366  | 90.5        | 305    | 2 E30838 | hemolysin E [impor |
| 4          | 1366  | 90.5        | 305    | 2 E85696 | probable pore form |
| 5          | 117   | 7.7         | 2139   | 2 T12896 | myosin heavy chain |
| 6          | 116.5 | 7.7         | 478    | 2 T12818 | hypothetical prote |
| 7          | 116.5 | 7.7         | 652    | 2 B59102 | hypothetical prote |
| 8          | 116   | 7.7         | 1023   | 1 LEECA  | hemolysin A - Esch |
| 9          | 115   | 7.6         | 1005   | 2 A64465 | hypothetical prote |
| 10         | 115   | 7.6         | 1093   | 2 AC1753 | tail protein [bact |
| 11         | 114.5 | 7.6         | 821    | 2 AC7087 | hypothetical prote |
| 12         | 113.5 | 7.5         | 587    | 2 A30394 | bp22 protein homol |
| 13         | 112.5 | 7.5         | 520    | 2 F70350 | recombination prot |
| 14         | 112.5 | 7.5         | 584    | 2 S75986 | hypothetical prote |
| 15         | 112.5 | 7.5         | 1024   | 2 S10056 | hemolysin A - Esch |
| 16         | 112.5 | 7.5         | 1098   | 2 B70232 | hypothetical prote |
| 17         | 112.5 | 7.5         | 2823   | 2 F87908 | protein T22A3.8 [i |
| 18         | 112.5 | 7.5         | 2823   | 2 T3064  | hypothetical prote |
| 19         | 112.5 | 7.5         | 3102   | 2 T43291 | laminin alpha chai |
| 20         | 111   | 7.4         | 329    | 1 VMT21  | VSG expression sit |
| 21         | 111   | 7.4         | 1295   | 2 T24587 | hypothetical prote |
| 22         | 110   | 7.3         | 1285   | 2 B72420 | hypothetical prote |
| 23         | 109   | 7.2         | 573    | 2 B97331 | membrane associate |
| 24         | 109   | 7.2         | 1492   | 2 T14652 | protein J - Yersin |
| 25         | 109   | 7.2         | 1545   | 2 T14966 | phage lambda-relat |
| 26         | 108.5 | 7.2         | 1959   | 1 A33977 | myosin heavy chain |
| 27         | 108.5 | 7.2         | 1999   | 1 S21801 | myosin heavy chain |
| 28         | 108   | 7.2         | 622    | 2 T22716 | hypothetical prote |
| 29         | 107   | 7.1         | 318    | 2 A45522 | variant surface gl |

|     |       |     |      |          |                    |
|-----|-------|-----|------|----------|--------------------|
| 30  | 106.5 | 7.1 | 1938 | 1 MWKWL  | myosin heavy chain |
| 31  | 106.5 | 7.1 | 2022 | 2 T43214 | ovti protein - nem |
| 32  | 106   | 7.0 | 1133 | 2 T22976 | hypothetical prote |
| 33  | 106   | 7.0 | 3187 | 2 JC5837 | 364K Golgi complex |
| 34  | 105   | 7.0 | 2405 | 2 T08164 | dynein alpha heavy |
| 35  | 104.5 | 6.9 | 1115 | 2 T41342 | probable coiled-co |
| 36  | 104.5 | 6.9 | 1292 | 2 D84727 | probable RAD50 DNA |
| 37  | 104.5 | 6.9 | 1964 | 2 A59282 | nonmuscle myosin I |
| 38  | 104   | 6.9 | 640  | 2 T03754 | hypothetical prote |
| 39  | 104   | 6.9 | 1496 | 2 T05634 | hypothetical prote |
| 40  | 104   | 6.9 | 2334 | 2 S32920 | cell wall-associat |
| 41  | 103.5 | 6.9 | 1163 | 2 G97236 | ATPase involved in |
| 42  | 103.5 | 6.9 | 4588 | 2 T28667 | dynein beta heavy  |
| 43  | 103   | 6.8 | 2017 | 1 A36014 | myosin II heavy ch |
| 44  | 103   | 6.8 | 2057 | 2 S61477 | myosin II heavy ch |
| 45  | 103   | 6.8 | 2155 | 2 AD2742 | conserved hypotet  |
| 46  | 103   | 6.8 | 2155 | 2 C97523 | hypothetical prote |
| 47  | 102.5 | 6.8 | 379  | 2 D97198 | methyl-accepting c |
| 48  | 102.5 | 6.8 | 472  | 2 H81665 | replicative DNA he |
| 49  | 102   | 6.8 | 726  | 2 T44825 | hypothetical prote |
| 50  | 102   | 6.8 | 739  | 2 H75001 | methyl-accepting c |
| 51  | 102   | 6.8 | 927  | 2 AG1739 | transmembrane prot |
| 52  | 102   | 6.8 | 1039 | 2 S18199 | myosin heavy chain |
| 53  | 102   | 6.8 | 1147 | 2 T40866 | cell polarity prot |
| 54  | 101.5 | 6.7 | 472  | 2 G71503 | probable replicati |
| 55  | 101.5 | 6.7 | 1127 | 2 T28317 | ORF MSV156 hypotet |
| 56  | 101.5 | 6.7 | 1473 | 2 A35186 | salivary agglutini |
| 57  | 101   | 6.7 | 1475 | 2 T33318 | hypothetical prote |
| 58  | 101   | 6.7 | 1601 | 2 AB1730 | hypothetical prote |
| 59  | 100.5 | 6.7 | 540  | 2 T44967 | gas-vesicle protei |
| 60  | 100.5 | 6.7 | 566  | 2 S54091 | hypothetical prote |
| 61  | 100.5 | 6.7 | 861  | 2 D82814 | ATP-dependent Clp  |
| 62  | 100.5 | 6.7 | 1009 | 2 B9792  | hypothetical prote |
| 63  | 100   | 6.6 | 451  | 2 T41722 | probable gamma-glu |
| 64  | 100   | 6.6 | 490  | 1 S71776 | calcium-dependent  |
| 65  | 100   | 6.6 | 779  | 2 E97778 | endopeptidase La ( |
| 66  | 100   | 6.6 | 927  | 2 AH1389 | transmembrane prot |
| 67  | 100   | 6.6 | 1278 | 2 T27925 | hypothetical prote |
| 68  | 100   | 6.6 | 1938 | 1 S06005 | myosin alpha heavy |
| 69  | 100   | 6.6 | 1938 | 2 T49464 | alpha cardiac myos |
| 70  | 99.5  | 6.6 | 579  | 2 G96966 | methyl-accepting c |
| 71  | 99.5  | 6.6 | 998  | 2 T00227 | hemolysin A toxin  |
| 72  | 99.5  | 6.6 | 1066 | 1 A48689 | kinesin-related pr |
| 73  | 99.5  | 6.6 | 1961 | 1 A61231 | myosin heavy chain |
| 74  | 99.5  | 6.6 | 2094 | 2 S33124 | tpz protein - huma |
| 75  | 99.5  | 6.6 | 2819 | 2 A90551 | conserved hypotet  |
| 76  | 99    | 6.6 | 481  | 2 B85575 | probable glutamate |
| 77  | 99    | 6.6 | 481  | 2 B90724 | probable glutamate |
| 78  | 99    | 6.6 | 520  | 2 S53393 | cell division cont |
| 79  | 99    | 6.6 | 539  | 2 F72288 | methyl-accepting c |
| 80  | 99    | 6.6 | 978  | 2 B89971 | conserved hypotet  |
| 81  | 99    | 6.6 | 1079 | 2 T18356 | membrane protein p |
| 82  | 99    | 6.6 | 1188 | 2 T18356 | chromosome segrega |
| 83  | 99    | 6.6 | 1322 | 2 B71440 | hypothetical prote |
| 84  | 99    | 6.6 | 2052 | 2 C97038 | phage-related prot |
| 85  | 99    | 6.6 | 2346 | 2 T13829 | tpz homolog - frui |
| 86  | 98.5  | 6.5 | 558  | 2 S62458 | vacuolar protein s |
| 87  | 98.5  | 6.5 | 826  | 2 T33796 | hypothetical prote |
| 88  | 98.5  | 6.5 | 1992 | 2 A47297 | myosin heavy chain |
| 89  | 98    | 6.5 | 401  | 2 E83720 | methyl-accepting c |
| 90  | 98    | 6.5 | 403  | 2 H72216 | outer membrane pro |
| 91  | 98    | 6.5 | 633  | 2 T41332 | casp homolog - fis |
| 92  | 98    | 6.5 | 916  | 2 T51288 | silencing protein  |
| 93  | 98    | 6.5 | 978  | 2 A70387 | conserved hypotet  |
| 94  | 98    | 6.5 | 1040 | 2 T40859 | silencing protein  |
| 95  | 98    | 6.5 | 1474 | 2 T18281 | hypothetical prote |
| 96  | 98    | 6.5 | 1939 | 1 A46762 | myosin alpha heavy |
| 97  | 97.5  | 6.5 | 760  | 2 C87029 | hypothetical prote |
| 98  | 97.5  | 6.5 | 804  | 2 B89961 | leucyl-tRNA synthe |
| 99  | 97.5  | 6.5 | 878  | 2 B84977 | alanine-tRNA ligas |
| 100 | 97.5  | 6.5 | 998  | 2 T41078 | hemolysin - Escher |

ALIGNMENTS

```
RESULT 1
AE0673
hemolysin HlyE [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain C
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AE0673
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AE0502; MUID:21534947; PMID:11677608
A:Accession: AE0673
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01758.1; PID:gl6502606; GSPDB:GN00176
C:Genetics:
C:Gene: STY1498
C:Superfamily: Escherichia coli hemolysin E

Query Match 97.5%; Score 1472; DB 2; Length 305;
Best Local Similarity 97.4%; Pred. No. 5.3e-91;
Matches 295; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MTGIFAEQTVVVKSAIETADGALDFYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 60
Db 3 MTGIFAEQTVVVKSAIETADGALDLYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 62

QY 61 LVGDIKVLMDSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 120
Db 63 LVGDIKVLMDSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 122

QY 121 DDGVNKLNEAQKSLGSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSQVDRIKKEAYAG 180
Db 123 DDGVNKLNEAQKSLGSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSQVDRIKKEAYAG 182

QY 181 AAAGIVAGPFGLLIISYIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVKQANKDIDAAK 240
Db 183 AAAGIVAGPFGLLIISYIAAGVIEGKLIPELNNRLKTVQNFTSLSVTVKQANKDIDAAK 242

QY 241 LKLAETIAAIGEIKETETTRFYVDYDDMLSLGAAKGMINTCNEYQQRHGKKTILEV 300
Db 243 LKLAETIAAIGEIKETETTRFYVDYDDMLSLGAAKGMINTCNEYQQRHGKKTILEV 302

QY 301 PDI 303
Db 303 PDV 305

RESULT 2
C64864
hemolysin E - Escherichia coli (strain K-12)
N:Alternate names: hemolysin-inducing protein
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: C64864
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64864
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-305 <BLAT>
A:Cross-references: GB:AE000216; GB:U00096; NID:gl1787417; PIDN:AACT4266.1; PID:gl1787430;
A:Experimental source: strain K-12, substrain MG1655
```

```
C:Genetics:
A:Gene: hlyE, hpr
C:Function:
A:Description: hemolytic activity
A:Note: pore formation
C:Superfamily: Escherichia coli hemolysin E
C:Keywords: cytolysis; cytotoxin; hemolysis; transmembrane protein
F:181-197/Domain: transmembrane #status predicted <TM>
F:123/Active site: Asp #status predicted

Query Match 91.5%; Score 1381; DB 2; Length 305;
Best Local Similarity 89.8%; Pred. No. 6.2e-95;
Matches 272; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

QY 1 MTGIFAEQTVVVKSAIETADGALDFYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 60
Db 3 MTEIVADKTVEVVKNAIETADGALDLYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 62

QY 61 LVGDIKVLMDSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 120
Db 63 LVGDIKVLMDSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 122

QY 121 DDGVNKLNEAQKSLGSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSQVDRIKKEAYAG 180
Db 123 DDGVNKLNEAQKSLGSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSQVDRIKKEAYAG 182

QY 181 AAAGIVAGPFGLLIISYIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVKQANKDIDAAK 240
Db 183 AAAGIVAGPFGLLIISYIAAGVIEGKLIPELNNKLSVQNFTSLSVTVKQANKDIDAAK 242

QY 241 LKLAETIAAIGEIKETETTRFYVDYDDMLSLGAAKGMINTCNEYQQRHGKKTILEV 300
Db 243 LKLAETIAAIGEIKETETTRFYVDYDDMLSLGAAKGMINTCNEYQQRHGKKTILEV 302

QY 301 PDI 303
Db 303 PEV 305

RESULT 3
E90838
hemolysin E [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90838
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genon
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90838
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <HAY>
A:Cross-references: UNIPROT:Q9REB3; GB:BA000007; PIDN:BA35100.1; PID:gl13361141; GSPDB:G
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
C:Gene: ECs1677
C:Superfamily: Escherichia coli hemolysin E

Query Match 90.5%; Score 1366; DB 2; Length 305;
Best Local Similarity 88.8%; Pred. No. 6.2e-84;
Matches 269; Conservative 20; Mismatches 14; Indels 0; Gaps 0;

QY 1 MTGIFAEQTVVVKSAIETADGALDFYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 60
Db 3 MTEIVADKTVEVVKNAIETADGALDLYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 62

QY 61 LVGDIKVLMDSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 120
Db 63 LVGDIKVLMDSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 122

QY 121 DDGVNKLNEAQKSLGSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSQVDRIKKEAYAG 180
```

[illegible]

**RESULT 4**  
E85696  
probable pore forming hemolysin hlyE [imported] - Escherichia coli (strain O157:H7, subspecies C; Species: Escherichia coli)  
C; Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C; Accession: E85696  
N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Perna, N.T.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
iller, L.  
ature 409, 529-533, 2001  
A; Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A; Reference number: A85480; UID:21074935; PMID:11206551

A:Accession: E85696  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-305 <STO>  
A:A:Cross-references: UNIPROT:Q9REB3; GB:AE005174; NID:gl2514879; PIDN:AAG56033.1; GSPDB:G  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: hlyE  
C:Superfamily: Escherichia coli hemolysin E

Query Match 90.5%; Score 1366; DB 2; Length 305;  
Best Local Similarity 88.8%; Pred. No. 6.2e-84;  
Matches 269; Conservative 20; Mismatches 14; Indels 0; Gaps 0

**QY** 1 MTGIFAEQTVVVKSAIETADGALDFYNKYLDQVIPWKTFTDETIKLSRFKEYSQEASV 60  
|||:::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
**DB** 3 MTEIVAKTVEVVKNAIETADGALDLYNKYLDQVIPWOTFDETIKLSRFKEYSQAASV 62

|    |    |   |     |
|----|----|---|-----|
| Qy | 61 | LVGDIKVLMDSDQKYFEATOTVYEWCGVTQLISAYILLFDEYNEKKASQAOKDILIRIL | 120 |
|    |    |   |     |
|    |    |   |     |
|    |    |   |     |
|    |    |   | :   |
|    |    |   | :   |
| dB | 63 | LVGNIKTLLMDSODKYFEATOTVYEWCGVATOLIAAYILLFDEYNEKKASQAOKLITKL | 122 |

|    |     |   |     |
|----|-----|---|-----|
| QY | 121 | DDGVNKLNEAQKSLGGSSQSFNNASGKLLALDSQITNDSEKSSVFQSQVDRIKRYAYAG | 180 |
|    |     | :   |     |
| b  | 123 | DDGITKLNPAQKSLIYSSQSFNNASGKLLALDSQITNDSEKSSVFQSQVDRIKRYAYAG | 182 |

[illegible]

QY           241 LKLAETAAIGETKTETTRFYVDYLMLSLKGAKMINTCNEYQQRHGKTKLLEV 300  
| | | | | | | | | | | | | | | | | | | | | | : |  
DB           242 TKTMTTAAICTATNTNTPNDPYNVNDI MISTEPMAMTMANVCPODUCYTPTREV 302

|    |             |
|----|-------------|
| Qy | 301 PDI 303 |
|    | ::          |
| Ch | 303 304 305 |

RESULT 5  
T18296  
myosin heavy chain - Entamoeba histolytica  
C:Species: Entamoeba histolytica  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18296  
R:Guillen, N.

submitted to the EMBL Data Library, February 1997

A;Reference number: Z18865

A;Accession: F18296

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-2139 <GUI>

A;Cross-references: UNIPROT:Q07569; EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AA84806

C;Genetics:

A;Gene: mhca

C;Superfamily: myosin heavy chain; myosin motor domain homology

F;91-780/Domain: myosin motor domain homology <WMO>

Query Match 7.7%; Score 117; DB 2; Length 2139;  
Best Local Similarity 19.0%; Pred. No. 15;  
Matches 75: Conservative 47; Mismatches 114; Indels 158; Gaps 13;

|    |     |   |      |
|----|-----|---|------|
| Qy | 7   | EQTVEVVKSAIETADGALDFYKNYLDQVWPKTFDETIKLSRFRQFYSQENASVLVGDIK | 66   |
| Db | 965 | EVETELNSOINTLNATVD-----KKDTIAEMOESIDEKEDETTKLGDIK           | 1011 |

|    |      |  |    |                    |      |
|----|------|--|----|--------------------|------|
| Qy |      | 67   | VL | -----LMDSODKYFEATQ | 81   |
|    | :    | :  | :  | :                  | :    |
| Db | 1012 | LLEERKDDLEODRADYSATKDIIAKINKITTECDRAKDEIAKLLOELEDENKNKDLTN |    |                    | 1071 |

|    |      |   |      |
|----|------|---|------|
| Qy | 82   | TVYEW--CGVVTQLLSAYILLFDEYNEKKASAOQDILIRILDGVGNKLN-----E         | 129  |
| Dh | 1072 | ELCOOTCKLGEFEKSTSAQVAAA-----TKKPSDEPTTSONI,EN--PKI,TTKNI,TKTKAD | 1124 |

QY           130 AQKSLIGSSQSF-----NNASGKLALDSQLTNDFFSKSSSYFQSQVDRIRK 175  
              |               |               |               |               |  
DB           1125 LEFKTSGLKNDYEDLDDKKNTGEGDIPNAGPKIKELIDDEITKGADVSQYLQVK-----K 1178  
              |               |               |               |               |

QY 176 BAYAGAAAGIVAGPGLIIISYSIAAGVIEGKLIPELNDRLKAVQNPFETSLSVTKQANKD 235  
          |              |              |              |              |              |  
DB 1178 EYFESQ                    IAYGGPPFAICNNUNKEVTIVSEIF 1212

QY 236 I-----DAALK--LATEIAAIGEITETETTRFYVDYDDLML----- 271

|    |      |                              |      |
|----|------|------------------------------|------|
| QY | 272  | -----SLLKGAAKKMINTCNEYQQRHGK | 294  |
| th | 1272 | -----SLLKGAAKKMINTCNEYQQRHGK | 1294 |

RESULT 6  
T12818

hypothetical protein yonD - *Bacillus subtilis* phage SPBC2  
C;Species: *Bacillus subtilis* phage SPBC2

C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
C;Accession: T12818; F69913

R; Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauer, C.; Karamata, D.

submitted to the EMBL Data Library, August 1997  
A;Description: The complete nucleotide sequence of the *Bacillus subtilis* SPbetac2 prophage  
A;Reference number: Z17583

A;Accession: T12818  
A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA  
A;Residues: 1-478 <L

A; Cross-references: UNIN  
P. Kunst F. : Ogasawara

Akman, S., O'Sullivan, M., Mozer, T., Liberman, M., Mazza, C., Mader, C., Bron, S., Brouillet, S., Bruschli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chofor, E.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390. 249-256. 1997

A.; Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funo, S.; Galizzi, A.; Gallert,  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;  
Koester, P.; Koningstein, G.; Kroch, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lindinois,  
A.; Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funo, S.; Galizzi, A.; Gallert,

A, Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Maueal, Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portercelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terptrapa, P.; Tognoni, A.; Toseato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: F69913  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-478 <KUN>  
A:Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14031.1; PID:el183560;  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yonD

Query Match 7.7%; Score 116.5; DB 2; Length 478;  
Best Local Similarity 23.0%; Pred. No. 2.3;  
Matches 58; Conservative 39; Mismatches 74; Indels 81; Gaps 12;

QY 25 DFYNYK-LDQVTPKTEFTETIKELSRFKQESQASVLVGIKVLMLDSQDKYPEATQTV 83  
Db 268 DVYDYFYVNVYSNS--DENSVD-KYFKNYTR-----TGDVTSIDPDKTEVF---MT 315  
QY 84 YEWCGVVTQLLSAYILLFDEYNEKKASAKQDILIRLDGYNKUNEAOKSLGSSQSPNN 143  
Db 316 RNWEEVWPEPIQSOL-----NQKDPQIKDLTKQVQNIN---KDKVGIEQQPNT 360  
QY 144 ASGKLALDSQ-----LNDPSEKSSYFQSQVDRIKRAYAGAAAGIVAG 188  
Db 361 ASEKLVQLNSEVQELKPYKHEKHTLLQKSEKNEFYKAFNALAE----- 409  
QY 189 PFGLIISYIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTKQANKDIDAAKLKLATEIA 248  
Db 410 -----KFS-----TEEVQNL---IHASVKQ-DEEGKAVQLNTMLV 442  
QY 249 AIGEIKTETETT 260  
Db 443 DLVSVPTETNTT 454

RESULT 7  
B59102  
hypothetical protein pX01-90 - *Bacillus anthracis* virulence plasmid pX01  
C:Species: *Bacillus anthracis*  
C:Date: 12-Nov-1999 #sequence\_revision 12-Nov-1999 #text\_change 09-Jul-2004  
A:Accession: B59102  
R:Okina, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler, J.  
J. Bacteriol. 181, 6509-6515, 1999  
A:Title: Sequence and organization of pX01, the large *Bacillus anthracis* plasmid harbored by *B. anthracis* strain Ames  
A:Reference number: A59091; MUID:99445483; PMID:10515943  
A:Accession: B59102  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-652 <OKI>  
A:Cross-references: UNIPROT:Q9X360; GB:AF065404; NID:g4894216; PIDN:AAD32394.1; PID:g4894216  
A:Experimental source: strain Sterne  
A:Note: similar to hypothetical protein; PPF0765w (980 aa); *Plasmodium falciparum* (AE001)  
C:Genetics:  
A:Gene: pX01-90  
A:Genome: plasmid

Query Match 7.7%; Score 116.5; DB 2; Length 652;  
Best Local Similarity 19.0%; Pred. No. 3.4;  
Matches 59; Conservative 65; Mismatches 106; Indels 81; Gaps 10;

QY 25 DFYNYKLDQVTPKTEFTETIKELSRFKQESQASVL-----VGIKVLMLDSQDKYPE 78  
Db 237 DVINOKIDE---FDKLSQRKDLERMLELNQKLSQIKQSPQLQDLNKKLXESQRLLE 293  
QY 79 ATQ-----TVYEWCGVVTQLLSAYILLFDEYNE--KKASAKQDILIRLDGYNK 126  
Db 294 LNKKDSNRNLNLSSEIKLNDKRAELSLIMELIKQSQSEFPDKIKNEKDLNKKREDLINR 353  
QY 127 LNEAQK-----SILGSSQFNNAAGKALDLSQNTDFSEKSYFQ---SOV 170  
Db 354 IASSELAKKAEALNTKLVELFKVQEAALNKSGQYLIIYINKLNLRELADKYKNSDNKI 413

QY 171 DRIRKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTVK 230  
Db 414 SRLKNH-----IGEVNKKLEKIE-----ELE 435  
QY 231 QANKDIDAARKLKLAT-EIAAIGEIKTETETTFYVDYDDL-----MLSLKGAAKMINT 284  
Db 436 ECNKKIDNTKQLABDFKSNKQOESELSVLQNKIDELGKRHKHQBLEASQKALDE 495  
QY 285 CNEYQORHGKK 295  
Db 496 AKBINKKLAEK 506

RESULT 8  
LEECA  
hemolysin A - *Escherichia coli*  
C:Species: *Escherichia coli*  
C:Date: 30-Sep-1988 #sequence\_revision 30-Sep-1988 #text\_change 09-Jul-2004  
C:Accession: A24433; I41280  
R:Palmllee, T.; Pellett, S.; Welch, R.A.  
J. Bacteriol. 163, 94-105, 1985  
A:Title: Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin.  
A:Reference number: A24433; MUID:85234404; PMID:3891743  
A:Accession: A24433  
A:Molecule type: DNA  
A:Residues: 1-1023 <PEL>  
A:Cross-references: UNIPROT:P09983; GB:M10133; GB:M12863; NID:gl46377; PIDN:AAA23975.1; I41280  
A:Experimental source: strain J96, O4 serotype  
R:Stanley, P.; Packman, L.C.; Koronakis, V.; Hughes, C.  
Science 266, 1992-1996, 1994  
A:Title: Fatty acylation of two internal lysine residues required for the toxic activity of hemolysin A  
A:Reference number: A55387; MUID:95093325; PMID:7801126  
A:Contents: annotation; lysine palmitoylation  
A:Note: lysine modification is performed by the hlyC gene product  
R:Haertlein, M.; Schiessi, S.; Wagner, W.; Rdest, U.; Kreft, J.; Goebel, W.  
J. Cell Biol. 22, 87-97, 1993  
A:Title: Transport of hemolysin by *Escherichia coli*.  
A:Reference number: I41280  
A:Accession: I41280  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1, 'T', '3', 'V', '5', 'T', '7-44 <RES>  
A:Cross-references: GB:M29173; NID:gl46337; PIDN:AAA23957.1; PID:gl46338  
C:Genetics:  
A:Gene: hlyA  
C:Function:  
A:Description: attacks blood cell membranes and causes cell lysis  
C:Superfamily: hemolysin A; hemolysin A homology  
C:Keywords: calcium binding; cytolysis; exotoxin; hemolysis; lipoprotein; tandem repeat;  
F:246-791/Domain: hemolysin A homology <HLYA>  
F:723-851/Region: 9-residue repeats [G-G-X-G-[DN]-D-X-[LVITYF]-X]  
F:563,685/Binding site: palmitate (Lys) (covalent) #status experimental

Query Match 7.7%; Score 116; DB 1; Length 1023;  
Best Local Similarity 20.2%; Pred. No. 6.7;  
Matches 68; Conservative 59; Mismatches 112; Indels 98; Gaps 11;

QY 2 TGITAEQTVEVVK-----SAIETADGALDFYKYLQVTPKTEFTETIKELSRFKQE 53  
Db 155 TALSSMKIDELIKKQSGNVSSSELAKASIELINQLVDTA-----ASLNNVNSFSQ 207  
QY 54 YSQBASVL-----VGIKVLMLDSQDKYPEATQTVYEWCGVVTQLLSAYILL-FDEY 104  
Db 208 LNKGLSVNTKHLNGVGN-KLQNLPLNDTGAGLDVT---SGILSAISAFILNADAD 263  
QY 105 NEKASAKQDILIRLDD---GVNKLEAOKSLGSSQSFNNASGKLIALDSQLTDFSE 161  
Db 264 TCTKAAGVELTTKVLGNVGKISQYIIAQRAAQGLSTS----- 302  
QY 162 KSSYFQSQVDRIKRAYAGAAAGIVAGPFGLIIS-----YSTAAGVIEGKLIPELNDRLKA 217  
Db 303 -----AAAAGLIASVVTLLAISPLSLSIADPKFRANKIEEYSQRPKK 344



478 LKSENEIDNKLKSLKETSLSKFNDS-----LIQINELISTKNNLQKQKMD----- 527  
 QY 177 AVAGAAAGIVAGPGLIISIAAGVIEGKLIPELN-DRLKAVQNFTSLSVTVKQAKND 235  
 Db 528 -----LNNLNDNLKVVQDKLIKNEETLKKEAE 556  
 QY 236 IDAAKL---KLATEIAA-----IGEKTE-----ETTRFYVDYDDL----- 269  
 Db 557 IDLSNSEMDLKKQTSKDDFKWQSKVEVEDEAKRNAEVTNLGNDIEDLKESKHL 616  
 QY 270 --MLSLKGAAKMINTCNEYQQRHGKTKLLEVPDI 303  
 Db 617 EETITELKNVHKLNEC-ELEKQKFETLSLESL 651

RESULT 12  
 A90394  
 bps2 protein homolog (bps2) [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
 C:Accession: A90394  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
 arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: A90394  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-587 <KUR>  
 A:Cross-references: UNIPROT:Q97MH8; GB:AE006641; NID:g13815540; PIDN:AAK42408.1; GSPDB:C  
 C:Genetics:  
 A:Gene: bps2

Query Match 7.5%; Score 113.5; DB 2; Length 587;  
 Best Local Similarity 19.5%; Pred. No. 4.7; Indels 87; Gaps 14;  
 Matches 64; Conservative 68; Mismatches 109; Indels 87; Gaps 14;  
 QY 3 GIPAEQVEVVK-----SAIETADGALDFYNYKLDQVPIKTFD 41  
 Db 10 GITQYSELEKGVTSYEAPNAVGTSLVRSLSLTSSTSIKAEDLLNVFADSGYVEAELD 69  
 QY 42 ETI--KELSRPKQYQSEASVLVDGDKVLLMSQDKYFEATQVYEWCGVVTOLLSA--- 96  
 Db 70 NKLYRRIRKIRNGLGEBKNLIMDDRALLL---TYFSFENR-----LVQILSGDGN 119  
 QY 97 ---YILLFDEYNEKKASQAKDILIRILDGYNKLNKAEQSLGSSOSFNN---ASQKLLA 150  
 Db 120 VEFVISTTSKINEIK-AKKELOKLLTABINARDELQK-----KYNINREIQAKIRA 170  
 QY 151 LDSQLTDFSEKSSYFQSQVDRIKAEYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPE 210  
 Db 171 ID-----BEIDKLEKE--RESSNIVAK-----TYTIT--LFRONKINE 206  
 QY 211 LNDRLKAVQNFTSLSVTVKQAKNDIDAAKLKLAETAAIGETETTRFYVDYDD-- 268  
 Db 207 ILNKIKVKKDELANLEFALKIEELQNKESKVSFDIKTQLEKEEINEKLLKLTNDRS 266  
 QY 269 ---LMLSLKGAKKM-----INTCN 286  
 Db 267 ELEIELKVLRLVEEVNESDRHLDTCN 294

RESULT 13  
 F70350  
 recombination protein RecN - Aquifex aeolicus  
 C:Species: Aquifex aeolicus  
 C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
 C:Accession: F70350  
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.

Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; MUID:98196666; PMID:9537320  
 A:Accession: F70350  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-520 <AQF>  
 A:Cross-references: UNIPROT:O66834; GB:AE000695; NID:g2983180; PIDN:AA06789.1; PID:g298;  
 A:Experimental source: strain VPS  
 C:Genetics:  
 A:Gene: recN  
 C:Superfamily: recN protein

Query Match 7.5%; Score 112.5; DB 2; Length 520;  
 Best Local Similarity 20.1%; Pred. No. 4.7;  
 Matches 59; Conservative 70; Mismatches 94; Indels 71; Gaps 13;  
 QY 10 VEVKSAIETADGALDFYNYKLDQVPIKTFD-----TIKELSRFKQESQ---EASVL- 61  
 Db 135 LEKVNSLRKKEQELFEFLRKEELIQRLEKVEEIGISSEVEELKNKANLIN 194  
 QY 62 -VGDIKVLMSQDKYFEATQVYEWCGVVTOLLS---AYILLFDEYNEKKASQAKDI-- 115  
 Db 195 NLEKVKAQGESLYKLEGENSVYELGEIRKNLAKVESYSGKFSSELIEKIANLEEVVE 254  
 QY 116 -----LIRILDGYNKLNKAEQSLGSSOSFNNASKLL---ALDSQLTN---DF 159  
 Db 255 LYNLSKEEMPEISEEVEINEKLPRIQRLEKYSKSPFELKEVEEIKKEELNSLNSVDF 314  
 QY 160 SEKSSYFQSQVDRIKAEYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELN-DRLKAV 218  
 Db 315 KEEE--LREEVEKLRBE-----YDKLAEVSRDRRKA 345  
 QY 219 QNFPTSLSVTVKQAKNDIDAAKLKLAETAAIGETETTRFYVDYDDLMS 272  
 Db 346 EDLEERIEELKEIN--LERAKLV-----EIK-ESEPTKYGDKIEFLFS 388

RESULT 14  
 S75986  
 hypothetical protein - Synechocystis sp. (strain PCC 6803)  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
 C:Accession: S75986  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 S:Reference number: S74322; MUID:97061201; PMID:8905231  
 A:Accession: S75986  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-584 <KAN>  
 A:Cross-references: UNIPROT:Q55486; EMBL:D64006; GB:AB001339; NID:g1001291; PIDN:BA01083;  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Superfamily: arginine-tRNA ligase

Query Match 7.5%; Score 112.5; DB 2; Length 584;  
 Best Local Similarity 23.2%; Pred. No. 5.4; Indels 75; Gaps 16;  
 Matches 76; Conservative 45; Mismatches 131; Indels 75; Gaps 16;  
 QY 3 GIPAEQVEVVKSAIETADGALD-----FY---NKYLDQVPIKTFETIKELSRFKQE 53  
 Db 171 GMLITYLKEVPEALVTAD-ALDIGLVTFYKQAKQRFQ-----DEQFRTSR----- 218  
 QY 54 YSOEASVLVGDIK-----VLMSQDKYFEATQVYEWCGVVTQ-----LISAVIL 99  
 Db 219 -QAVVALQAGDAKSIAKAWQLCEQSRREF---QLIYDCLDITIERGCSFYNPFLPGVVE 274  
 QY 100 LFDEYNEKKASQAKDILIR-----ILDDGYNKLNKAEQSLGSSQ-SFNNASKLIAL 151





Wed Feb 2 11:26:41 2005

us-09-993-292b-24.rpr

C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.5%; Score 112.5; DB 2; Length 2823;  
Best Local Similarity 19.9%; Pred. No. 44;  
Matches 73; Conservative 64; Mismatches 106; Indels 123; Gaps 16;

QY 6 ABQTVVV-----KSAIETADGALDFYKYLQVDPKTFDETIKELSRFKQYEQAS 59  
DB 1862 ABELKMWTAQKLNITFEDLNKRDVLEQWMD-----YRETIYDVSKDTADAERMS 1915

QY 60 VLVG-----DIKVLMDSDQK-----YFEATQTV----- 83  
DB 1916 LVVGRKRINRYKEVSEIEKLRVEADQIAYSRNSIEKARSEELMNMFDKEKINMTLAE 1975

QY 84 ----YEWCGVVTQLLSAYILLFDEYNEK-----KASAKDILIRILDGVNKLN 128  
DB 1976 PDLVEQCONITLL-----YSOLIDEYDEEVVQTAGRAHAEKLEVQAOK-IVDRFVDTRTETEN 2031

QY 129 -----BAOKS-----LLGSSQSFNNASGKLLALDSQLTNDPSEKSYFOSQVDRI 173  
DB 2032 PLKASHAYENIVEALKNATEAVDSAAEASEAVSKMLGSESGSDANEES--LRSOLEKL 2089

QY 174 RKEAYAGAAAGIVAGPFGILLIISYIAAGVIE--GKLIPELNDRLKAVQNFETSLSVTVKQ 231  
DB 2090 KNESSLSN-----VDNSNAVKIVEELKKEKDLTDLRGLHNLKTSI----- 2131

QY 232 ANKIDAAKLKLAETEIAAIGETETETTRFVVDYDDLMLSLKGAACKM-INTCNEYQQ 290  
DB 2132 -----VKRLGVIKNEASS-----WDDKHDRMHSILKNGAKTAHERSANVKE 2173

QY 291 RHGKKT 296  
DB 2174 SEGIKT 2179

RESULT 18

T23064  
hypotheical protein T22A3.8 - Caenorhabditis elegans (fragment)  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T23064; T25096  
R:Barlow, K.  
submitted to the EMBL Data Library, October 1997  
A:Reference number: Z19669  
A:Accession: T23064  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2823 <W12>  
A:Cross-references: UNIPROT:O45614; EMBL:AL008585; PIDN:CAA15432.1; GSPDB:GN00019; CESP:  
A:Experimental source: clone H10E24  
R:McMurray, A.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19980  
A:Accession: T25096  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2823 <W12>  
A:Cross-references: EMBL:281125; PIDN:CAB03385.1; GSPDB:GN00019; CESP:T22A3.8  
A:Experimental source: clone T22A3  
C:Genetics:  
A:Gene: CESP:T22A3.8  
A:Map position: 1  
A:Introns: 45/1; 282/2; 312/3; 416/2; 1255/3; 1329/3; 1418/3; 1776/2; 1988/2; 2760/2  
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.5%; Score 112.5; DB 2; Length 2823;  
Best Local Similarity 19.9%; Pred. No. 44;  
Matches 73; Conservative 64; Mismatches 106; Indels 123; Gaps 16;

QY 6 ABQTVVV-----KSAIETADGALDFYKYLQVDPKTFDETIKELSRFKQYEQAS 59  
DB 1862 ABELKMWTAQKLNITFEDLNKRDVLEQWMD-----YRETIYDVSKDTADAERMS 1915

QY 60 VLVG-----DIKVLMDSDQK-----YFEATQTV----- 83  
DB 1916 LVVGRKRINRYKEVSEIEKLRVEADQIAYSRNSIEKARSEELMNMFDKEKINMTLAE 1975

QY 84 ----YEWCGVVTQLLSAYILLFDEYNEK-----KASAKDILIRILDGVNKLN 128  
DB 1976 PDLVEQCONITLL-----YSOLIDEYDEEVVQTAGRAHAEKLEVQAOK-IVDRFVDTRTETEN 2031

QY 129 -----BAOKS-----LLGSSQSFNNASGKLLALDSQLTNDPSEKSYFOSQVDRI 173  
DB 2032 PLKASHAYENIVEALKNATEAVDSAAEASEAVSKMLGSESGSDANEES--LRSOLEKL 2089

QY 174 RKEAYAGAAAGIVAGPFGILLIISYIAAGVIE--GKLIPELNDRLKAVQNFETSLSVTVKQ 231  
DB 2090 KNESSLSN-----VDNSNAVKIVEELKKEKDLTDLRGLHNLKTSI----- 2131

QY 232 ANKIDAAKLKLAETEIAAIGETETETTRFVVDYDDLMLSLKGAACKM-INTCNEYQQ 290  
DB 2132 -----VKRLGVIKNEASS-----WDDKHDRMHSILKNGAKTAHERSANVKE 2173

QY 291 RHGKKT 296  
DB 2174 SEGIKT 2179

RESULT 19

T43291  
laminin alpha chain - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T43291  
R:Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang,  
submitted to the EMBL Data Library, June 1998  
A:Description: Expression, function and evolution of laminin alpha chains.  
A:Reference number: Z22397  
A:Accession: T43291  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-3102 <ZHU>  
A:Cross-references: UNIPROT:O45614; EMBL:AF074902; PIDN:AAC26793.1  
C:Genetics:  
A:Map position: 1  
A:Note: lamal/2  
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.5%; Score 112.5; DB 2; Length 3102;  
Best Local Similarity 19.9%; Pred. No. 50;  
Matches 73; Conservative 64; Mismatches 106; Indels 123; Gaps 16;

QY 6 ABQTVVV-----KSAIETADGALDFYKYLQVDPKTFDETIKELSRFKQYEQAS 59  
DB 1862 ABELKMWTAQKLNITFEDLNKRDVLEQWMD-----YRETIYDVSKDTADAERMS 1915

QY 60 VLVG-----DIKVLMDSDQK-----YFEATQTV----- 83  
DB 1916 LVVGRKRINRYKEVSEIEKLRVEADQIAYSRNSIEKARSEELMNMFDKEKINMTLAE 1975

QY 84 ----YEWCGVVTQLLSAYILLFDEYNEK-----KASAKDILIRILDGVNKLN 128  
DB 1976 PDLVEQCONITLL-----YSOLIDEYDEEVVQTAGRAHAEKLEVQAOK-IVDRFVDTRTETEN 2031

QY 129 -----BAOKS-----LLGSSQSFNNASGKLLALDSQLTNDPSEKSYFOSQVDRI 173  
DB 2032 PLKASHAYENIVEALKNATEAVDSAAEASEAVSKMLGSESGSDANEES--LRSOLEKL 2089

QY 174 RKEAYAGAAAGIVAGPFGILLIISYIAAGVIE--GKLIPELNDRLKAVQNFETSLSVTVKQ 231  
DB 2090 KNESSLSN-----VDNSNAVKIVEELKKEKDLTDLRGLHNLKTSI----- 2131

QY 232 ANKIDAAKLKLAETEIAAIGETETETTRFVVDYDDLMLSLKGAACKM-INTCNEYQQ 290  
DB 2132 -----VKRLGVIKNEASS-----WDDKHDRMHSILKNGAKTAHERSANVKE 2173

QY 291 RHGKKT 296  
DB 2174 SEGKKT 2179

RESULT 20  
NMUT21  
VSG expression site-associated protein 221a precursor - Trypanosoma brucei  
N:Alternate names: ESAG protein  
C:Species: Trypanosoma brucei  
C>Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 09-Jul-2004  
C:Accession: A03395  
R:Cully, D.F.; Ip, H.S.; Cross, G.A.M.  
Cell 42, 173-182, 1985  
A:Title: Coordinate transcription of variant surface glycoprotein genes and an expression site-associated protein  
A:Reference number: A90868; MUID:85254917; PMID:2861910  
A:Accession: A03395  
A:Molecule type: mRNA  
A:Residues: 1-329 <CUL>  
A:Cross-references: UNIPROT:P04478; GB:M11452; NID:G162072; PIDN:AAA30191.1; PID:G162073  
C:Comment: The function of the ESAG proteins is not known but may be related to activation of the VSG expression site  
C:Superfamily: VSG expression-site associated protein  
C:Keywords: glycoprotein  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-329/Product: VSG expression site-associated protein 221a #status predicted <MAT>  
F:73,294,308/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 7.4%; Score 111; DB 1; Length 329;  
Best Local Similarity 22.9%; Pred. No. 3.2;  
Matches 48; Conservative 33; Mismatches 63; Indels 56; Gaps 9;

QY 92 QLLSAYILLFD-----EYNEKKASAKDILIRILDGVNKL-NEAOKS 133  
DB 4 EIVELVLLFVTCVDWMLQAGDCTRVADHKHAPTEAVCVLRCLSDALNKLKLYGEKK 63

QY 134 LLGSSQSFNNAAGKLLALDSQTNDFSEKSY-----FQSQVDRIKREYAGAAAGIV 186  
DB 64 LLVTEEVANAS---LILD-DMEGRAGESSTVSLVIRGVMBEQTDRLK-----108

QY 187 AGPFLGIISYSIAAGVI---EGKLPELNDRLKAVQ-----NFFTSLSVTVKQA 232  
DB 109 -----LISYGNMGVNAKAGGLFAALDSLSKVRKIPGALIKTNKYYSVAEIVRTV 162

QY 233 NKDIDAALKL-----ATEIAAIGEIKTE 256  
DB 163 WEDVGEILWKETEAKGSKQVGVGEIQTE 192

RESULT 21  
T24587  
hypothetical protein T06E4.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T24587  
R:Lloyd, C.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: T24587  
A:Accession: T24587  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1295 <WIL>  
A:Cross-references: UNIPROT:Q22257; EMBL:270756; PIDN:CAA94789.1; GSPDB:GN00023; CESP:T06E4  
A:Experimental source: clone T06E4  
C:Genetics:  
A:Gene: CESP:T06E4.1  
A:Map position: 5  
A:Introns: 17/3; 218/2; 385/3; 481/3; 946/3; 1034/3; 1222/2

Query Match 7.4%; Score 111; DB 2; Length 1295;  
Best Local Similarity 21.0%; Pred. No. 20;  
Matches 65; Conservative 58; Mismatches 93; Indels 94; Gaps 13;

QY 6 AEQTVVVKSAIETADGALDFYNYKLDQVIPWKTFTET-----IKLSRFKQYSQASVL 61

DB 710 AEETNEKLSDLASSE-----EQILDKNQQLSDLDLKEKLSAESTNQELQVS 759  
QY 62 VGDIKVLLMDSQDKYFEATQTVYEWGCVVTOILLSAYILLFDEYNEKKASA-----OKD1 115  
DB 760 LEMLKIEVSNARQKWESE-----VLKESFEALQLELSASQVSRVVDAAVQEKDG 811

QY 116 LIRILDDGVNKLNEAOKS---LLGSS-----QSFN---NASGK- 147  
DB 812 LLRLVDTEKLKIEDTEKSAQDLQOSSVBEIKQLQLDLQNFQNAEVLSEKLNKLNSSHKR 871

QY 148 -LLALDSQTNDFSEKSYFQSQVDRIKREYAGAAAGIVAGPFLGIISYSIAAGVIEGK 206  
DB 872 DMVALASQL-BELQHKLVVGSQVENVKEEL-----IGAKIMNKE 910

QY 207 LIPELN---DRLKAVONFFTSLSVT-----VKQANKDIDAAKLKLAIEAIG 251  
DB 911 MVDELNAKLGDALEGMELKSLSEVSEAKVORREELIAQVSKHRDQES-QLQTLDELK 969

QY 252 EIKTETETTR 261  
DB 970 SAQHSSTETSR 979

RESULT 22  
B72420  
hypothetical protein TM0088 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: B72420  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: B72420  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1285 <ARN>  
A:Cross-references: UNIPROT:Q9WXU3; GB:AE001695; GB:AE000512; NID:G4980569; PIDN:AAD35181  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0088

Query Match 7.3%; Score 110; DB 2; Length 1285;  
Best Local Similarity 19.5%; Pred. No. 23;  
Matches 69; Conservative 69; Mismatches 113; Indels 102; Gaps 20;

QY 1 MTGIFAEQT-----VEVVKSAIETADGALDFYNYKLDQ-----VIPWKTFTETIKELGRF 50  
DB 487 MYGVSMKIGNFLEFGARSSVENVEQL---KKLSBHTQFLKALKT--ENVEDLGRF 541

QY 51 -KQYEQASVILVDGIKVLMDSQDKYFEATQTVYEWGCVVTOILLSAYILLFDEYNEK-K 108  
DB 542 MKKYGVEFEYF--SSLKVMAMLSGKEE--ENVQKAABELQIISS--EERIIFVKKTENVP 596

QY 109 ASAKDILIRILDDGVNKLNEAOKSLLGSSQSFNNAAG---KLLALDSQTNDFSEKSSY 165  
DB 597 IDKAKNVVLQYSYSIEELGN-ELVVIGEREVEKADLLQKIFSEVEISRDFVKLPSW 655

QY 166 FQSQVDRIKREYAGAAAGIVAGPFLGIISYSIAAGVI--EGKLPELNDRLKAVONFFT 223  
DB 656 IDEQEKLE-----VVKNSAG--ITVEILDGVVYFEG-----685

QY 224 SLSVTVQKANKDIDAAKLKLAIEAIGETETETTRF-----YV 264  
DB 686 -----TKENVEKAKELFSDIVEKLGEVRKB-ETVEFLEVNSGFFVDEFINLSGKLYP 736

QY 265 D-----YDDLMLSLKLGAAKMKMINTCNEYQ---ORHGK-----KTLLEVP 301  
DB 737 DVTCSLDQLGLLVKGSSEAVEDLSSMYRFFERHQRKIVKENVFDRLMLEVP 789



A:Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular my  
A:Reference number: A33977; MUID:90046668; PMID:2813355  
A:Accession: A33977  
A:Molecule type: mRNA  
A:Residues: 1-1959 <SHO>  
A:CROSS-references: UNIPROT:P14105; GB:M26510; NID:9212382; PIDN:AAA48974.1; PID:9212383  
R:Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.  
Eur. J. Biochem. 184, 611-616, 1989  
A:Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed in  
s.  
A:Reference number: S06116; MUID:90032648; PMID:2806244  
A:Accession: S06116  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 716-1008 <KAT>  
A:CROSS-references: GB:X17589  
A:Note: this translation is not annotated in GenBank entry GGMCFMHA, release 114  
R:Hodge, T.P.; Cross, R.; Kendrick-Jones, J.  
J. Cell Biol. 118, 1085-1095, 1992  
A:Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate  
A:Reference number: A43422; MUID:92381096; PMID:1512291  
A:Accession: A43422  
A>Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1900-1959 <HOD>  
A:Experimental source: brush border  
A:Note: sequence extracted from NCBI backbone (NCBIP:111947)  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide  
F:84-764/Domain: myosin motor domain homology <MMOT>  
F:174-181/Region: nucleotide-binding motif A (P-loop)  
F:552-565/Region: actin binding #status predicted  
F:626-640/Region: actin binding #status predicted  
F:837-1936/Domain: coiled coil #status predicted <COI>  
F:837-1277/Region: S2  
F:1278-1959/Region: light meromyosin  
F:1937-1959/Domain: carboxyl-terminal <CBT>  
F:125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted  
F:180/Binding site: ATP (Lys) #status predicted  
F:694,704/Active site: Cys #status predicted

Query Match 7.2%; Score 108.5; DB 1; Length 1959;  
Best Local Similarity 21.0%; Pred. No. 50;  
Matches 66; Conservative 43; Mismatches 107; Indels 99; Gaps 11;  
QY 7 EQTVVVKSAIB---TADGALDFNKKYLDQVIPWKTDETIKE-----LSRFKQEYS 55  
DB 1166 EQEVTVLKTLEDEAKTHEAQIQEMRQKHSQAI--EELAEQLEQTKRVKANLEKAKQALE 1223  
QY 56 QEASVLVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLDFYNEKKASAKQDI 115  
DB 1224 SERAELSEVKKVLLQKGD-----EHRKKVDAQLQE 1256  
QY 116 L-----IRLLDGVNKLNEAKSLGSSQSFNNASGKLLALDS---QLTNDFSEKSY 165  
DB 1257 LQVKTPEGRVKTLEAERVKNLQVEL-----DNVTGLNQSDSKSIKLAKDFSALESQ 1309  
QY 166 FQSQVDRIKRAYAGAAAGIVAGPGLIISYSIAAGVTEGK---LIPELNDRLKAVQNF 222  
DB 1310 LQDTQLQDET-----RLKLSFTKLTQTEDEKNAKLEQLEEEEAERMLE 1356  
QY 223 TSLSVTVKQA-----NKDIDAAKLKATIAAIGEIKTETE 258  
DB 1357 KQISVLVQQAQVARKKMDGLGCLAEAEAKKQLKDLLESLTQRYEEKIAAYD--KLEKT 1414  
QY 259 TTRFVVDVDDMLSL 273  
DB 1415 KTRLQQLDLDIAVDL 1429

RESULT 27

S21801  
myosin heavy chain, neuronal [similarity] - rat

N:Alternate names: myosin II  
N:Contains: myosin ATPase (EC 3.6.4.1)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C:Accession: S21801; PN0013; S18134  
R:Sun, W.; Chantler, P.D.  
J. Mol. Biol. 224, 1185-1193, 1992  
A:Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain  
A:Reference number: S21801; MUID:92235856; PMID:1569576  
A:Accession: S21801  
A:Molecule type: mRNA  
A:Residues: 1-1999 <SUN>  
A:CROSS-references: UNIPROT:Q63731; EMBL:X62659  
R:Sun, W.; Chantler, P.D.  
Biochem. Biophys. Res. Commun. 175, 244-249, 1991  
A:Title: A unique cellular myosin II exhibiting differential expression in the cerebral c  
A:Reference number: PN0013; MUID:91151356; PMID:1998509  
A:Accession: PN0013  
A:Molecule type: mRNA  
A:Residues: 1914-1998, 'I' <SU2>  
A:Experimental source: brain  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide  
F:84-763/Domain: myosin motor domain homology <MMOT>  
F:174-181/Region: nucleotide-binding motif A (P-loop)  
F:541-575/Region: actin binding #status predicted  
F:653-675/Region: actin binding #status predicted  
F:836-1999/Domain: coiled coil #status predicted <COI>  
F:836-1276/Region: S2  
F:1277-1999/Region: light meromyosin  
F:125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted  
F:180/Binding site: ATP (Lys) #status predicted  
F:693,703/Active site: Cys #status predicted  
F:1916/Binding site: phosphate (Ser) (covalent) #status predicted  
F:1943/Binding site: phosphate (Ser) (covalent) #status predicted  
Query Match 7.2%; Score 108.5; DB 1; Length 1999;  
Best Local Similarity 19.5%; Pred. No. 51;  
Matches 66; Conservative 53; Mismatches 124; Indels 95; Gaps 13;  
QY 7 EQTVVVKSAIE---TADGALDFNKKYLDQVIP--WKTDEET---IKELSRFKQEYSQE 57  
DB 1165 EQEVNLIKTLLEERAKTHEAQIQEMRQKHSQAVEELASQLEQTKRVKANLEKAKOTLENE 1224  
QY 58 ASVLVGDITKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLDFYNEKKASAKQDILI 117  
DB 1225 RGLANEVKKVLLQGRD-----SEHKKRVKVEAQLQELQ 1257  
QY 118 RILDDGVNKLNEAKSLGSSQSFNNASGKLLALD---SOLTNDPS----- 160  
DB 1258 VKFNEGERRVTELADKVTKLQVELDNVTGLLSQSDSKSKLTKDFSALESQLDQTELLQ 1317  
QY 161 -----EKSSYFQSOVDRIKRAYAGAAAGIVAGPGLIISYSIAAGV 202  
DB 1318 EENRQKLSLTKLQVDEKNS--FREQLLEEEEAERMLEKQIA-----TLHAQV 1366  
QY 203 IEGKLIPELNDRLKAVQNFFTSLSVTVKQANKDDAAKLKATIAAIGEIKTETETTRF 262  
DB 1367 ADMK--KQEDSVGCL-----TAEVKKKQKDLGLESLQRHEEKVAYD--KLEKTKRL 1418  
QY 263 YVDVDDMLSLKGAQKQWINTCN--EYQQRHGKTKTLE 299  
DB 1419 QOELDDLLVDL----DHQRQSACNLKKKQKQKFDQLLAE 1452

RESULT 28

T27216

hypothetical protein F55C5.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T27216

R:Harris, B.

submitted to the EMBL Data Library, August 1996

Db 136 --DALMRANKYYTAIAEVVTVWDDVKAVE 163

RESULT 30

MYKWI

myosin heavy chain D [similarity] - Caenorhabditis elegans

N;Alternate names: myosin heavy chain 1

N;Contains: myosin ATPase (EC 3.6.4.1)

C;Species: Caenorhabditis elegans

C;Date: 28-Feb-1986 #sequence revision 19-May-2000 #text\_change 09-Jul-2004

C;Accession: T21193; T23973; S02772; A02993

R;McMurray, A.

submitted to the EMBL Data Library, April 1996

A;Reference number: Z19388

A;Accession: T21193

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1938 <WIL>

A;Cross-references: UNIPROT:P02567; EMBL:Z71261; PIDN:CAA95806.1; GSPDB:GN00019; CESP:R06

A;Experimental source: clone F21C3

R;Gardner, A.

submitted to the EMBL Data Library, April 1996

A;Reference number: Z19825

A;Accession: T23973

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1938 <W12>

A;Cross-references: EMBL:Z71266; PIDN:CAA95848.1; GSPDB:GN00019; CESP:R06C7.10

A;Experimental source: clone R06C7

R;Dibb, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.

J. Mol. Biol. 205, 603-613, 1989

A;Title: Sequence analysis of the complete Caenorhabditis elegans myosin heavy chain gene

A;Reference number: S02771; MUID:89178677; PMID:2326820

A;Accession: S02772

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-376, 'V', 378-390, 'V', 392-576, 'L', 578-680, 'I', 682-1938 <DIB>

A;Cross-references: EMBL:X08065; NID:g6785; PIDN:CAA30854.1; PID:g6786

R;Karn, J.; Brenner, S.; Barnett, L.

Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983

A;Title: Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy chain

A;Reference number: A93958; MUID:83273600; PMID:6576334

A;Accession: A02993

A;Molecule type: DNA

A;Residues: 24-93, 'E', 95-97, 'R', 99-376, 'V', 378-388, 'GDV', 392-407, 'N', 409-473, 'G', 475-576, 'C';

C;Genetics:

A;Gene: myo-1; CESP:R06C7.10

A;Map position: 1

A;Introns: 23/3; 114/3; 229/1; 264/1; 320/1; 857/3; 1745/3; 1814/1; 1892/3

C;Superfamily: myosin heavy chain; myosin motor domain homology

C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle cor

F;87-773/Domain: myosin motor domain homology <MMOT>

F;177-184/Region: nucleotide-binding motif R (P-loop)

F;660-682/Region: actin binding #status predicted

F;764-778/Region: actin binding #status predicted

F;846-1938/Domain: coiled coil #status predicted <COI>

F;846-1160/Region: S2

F;1161-1938/Region: light meromyosin

F;128/Modified site: N6,N6-trimethyllysine (Lys) #status predicted

F;183/Binding site: ATP (Lys) #status predicted

F;700,710/Active site: Cys #status predicted

Query Match 7.1%; Score 106.5; DB 1; Length 1938;

Best Local Similarity 17.3%; Pred. No. 67;

Matches 52; Conservative 63; Mismatches 112; Indels 73; Gaps 8;

QY 5 FAEQTVVVK-----SAIETADGALDPYKYLDOVPWTFDETIKELSRFK 51

Db 1315 FSSQLVAAKAAEDELHERQEFHAAKNEHLEHDOCHELEEQINGK--DIQRLSRIN 1372

QY 52 QEVSQ-----EASLVG--DIKVLMDSQDKFEATQTVYEWCGVVTQLLSAYILLFDEY 104

Db 1373 SEISQWKARYEGEGLVGSELEELKQKNRVMIDLQEALSAQNKVLSLEAKGKLLAET 1432

A;Reference number: Z19604

A;Accession: T22716

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-622 <WIL>

A;Cross-references: UNIPROT:Q02082; EMBL:Z78198; PIDN:CAB01573.1; GSPDB:GN00023; CESP:R06

A;Experimental source: clone F55C5

C;Genetics:

A;Gene: CESP:F55C5.8

A;Map position: 5

A;Introns: 24/3; 44/2; 147/3; 251/3; 561/3

Query Match 7.2%; Score 108; DB 2; Length 622;

Best Local Similarity 20.6%; Pred. No. 12;

Matches 65; Conservative 54; Mismatches 111; Indels 86; Gaps 14;

QY 41 DFTIKELSRFKQYEQEASVLVGDILKVLMDSQDKYFEATQTVYEWCGVVTQLLSAY--- 97

Db 251 DKLISEM-RASATSAEVVTVIEWGGAKSVDD-----KAKQVQEWKQTEVELAQCTPK 304

QY 98 --ILLFDEYNEKKAQAKDILRIILDDGVNKLNE-AQKSLIGSSQSF-----NNASGKL 148

Db 305 EKWALF---ERATADTRDAIDRISDIIRKSSSENADITVLQSIKAYLEFLKMGTSARY 360

QY 149 LALDSQLTNDFSEKSSYFQS-----QVDRIRKEAYAG 180

Db 361 LAI---IDNTSEKSKPODLLRLVDSVIEIYKEVAEIPGADHKVLQAEVKEVYRA 417

QY 181 AAGIVAGPFGILISIAAGVI-----EGKLPELMDRLKAVQNFSTLSVTVK 230

Db 418 FRCFYNASYSALHKYSEAAALFDRTVSRVQDAEGKL-----KLLKSSSFITNETQSSIN 472

QY 231 QANKDIDAAK--IKLATEATAAIGETETTFRFYVDYDMLSLKKAAGKMGINTCNEY 288

Db 473 ELRSEVESAKTVRAARLASAGDVKTDSLAKI-ID-----KRELLETVNEW 519

QY 289 QO---RHGKTKLLEVP 301

Db 520 QWDVNRSLKDKKTIIP 535

RESULT 29

A45522

variant surface glycoprotein U1EC - Trypanosoma brucei gambiense (fragment)

C;Species: Trypanosoma brucei gambiense

C;Date: 03-Jun-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004

C;Accession: A45522

R;Barnes, D.A.; Motttram, J.C.; Agabian, N.

Mol. Biochem. Parasitol. 41, 101-114, 1990

A;Title: Bloodstream and metacyclic variant surface glycoprotein gene expression sites

A;Reference number: A45522; MUID:90348709; PMID:2385264

A;Accession: A45522

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-318 <BAR>

A;Cross-references: UNIPROT:Q26749; GB:M31612; NID:gl62074; PIDN:AAA30192.1; PID:gl62075

C;Superfamily: VSG expression-site associated protein

C;Keywords: Glycoprotein

Query Match 7.1%; Score 107; DB 2; Length 318;

Best Local Similarity 27.3%; Pred. No. 5.6;

Matches 41; Conservative 27; Mismatches 64; Indels 18; Gaps 8;

QY 100 LPDEYNEKKAQAKDILRIILDDGVNKL-NFAQKSLIGSSQSFNNASGKLALDSQLTND 158

Db 23 LVDDYIGKQLH-ESVCYLRLCSNALNKLYSGEKKLLATEEVVANAS---LILD-DMGR 77

QY 159 FSEKSSY-----FQSQVDRIRK-EAYAGAAAGIVAGPGLIISYIAAGVIEGKLPIE 210

Db 78 AGESSTYLSIISYGVMEQTDRLKLIISYGNAMGDLVAKAGLFAEVNESVRAVR-KKIP- 135

QY 211 LNDRLKAVQNFSTLSVTVKQANKDIDAAK 240

```
QY 105 NEKASAKDILII-----RIILDDGVNKLNEAQKSLGSSQFNNASGKLALD 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1433 EDARSDVRHLTVIASLEKKQAFKIVDDMKRKVDIIQKEIDATRDSTRNSTEVEFKL 1492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 153 SOLTNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELN 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1493 SSMND-----LSEQIETLRE-----NKIFSQSIRIDN 1520
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 213 DRLKAVQNFFTSLSVTVQANKD-----IDAAKCLKLATEAAGIETETTTTFYVD 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1521 EOTGOGTYQEVHKSVRLEQKDELOHALDEAALAEBSKVRLQIEVQQRSEIE 1580
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 31
T43214
Nv1 protein - nematode (Onchocerca volvulus)
A:Alternate names: myosin-like antigen
C:Species: Onchocerca volvulus
C:Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T43214; A44939; A54513; S27825
R:Triteerapapab, S.; Richie, T.L.; Tuan, R.S.; Shepley, K.J.; Dinman, J.D.; Neubert, T.
Mol. Biochem. Parasitol. 69, 161-171, 1995
A:Title: Molecular cloning of a gene expressed during early embryonic development in Onchocerca
A:Reference number: Z22341; MUID:95287898; PMID:7770081
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Accession: T43214
A:Molecule type: mRNA
A:Residues: 1-2022 <TRI>
A:Cross-references: UNIPROT:P21249; EMBL:U12691; NID:G530824; PID:G530825; PIDN:AAA80009
A:Experimental source: specific_host Homo sapiens
R:Erondou, N.E.; Donelson, J.E.
Mol. Biochem. Parasitol. 40, 213-224, 1990
A:Title: Characterization of a myosin-like antigen from Onchocerca volvulus.
A:Reference number: A44939; MUID:90301142; PMID:2194123
A:Accession: A44939
A:Molecule type: mRNA
A:Residues: 733-874, 'E', 876-916, 'S', 918-1038, 1040-1047, 'S', 1049-1283, 'E', 1285-1363 <ERO>
A:Cross-references: GB:M30398
A:Note: the sequence is revised in GenBank entry ONGANTWL, release 115, (PIDN:AAA29413.1)
R:Donelson, J.E.; Duke, B.O.L.; Moser, D.; Zeng, W.; Erondou, N.E.; Lucius, R.; Renz, A.
Mol. Biochem. Parasitol. 31, 241-250, 1988
A:Title: Construction of Onchocerca volvulus cDNA libraries and partial characterization
A:Reference number: A54513; MUID:89127417; PMID:2464764
A:Accession: A54513
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 733-866 <DON>
A:Cross-references: GB:J03995; NID:G159874; PIDN:AAA29412.1; PID:G159875
R:Ngozi, E.; Erondou, N.E.; Donelson, J.E.
submitted to the EMBL Data Library, April 1990
A:Description: Characterization of a myosin-like antigen from Onchocerca volvulus.
A:Reference number: S27825
A:Accession: S27825
A:Molecule type: mRNA
A:Residues: 733-874, 'E', 876-916, 'S', 918-1038, 1040-1047, 'S', 1049-1283, 'E', 1285-1347, 'E', 1
A:Cross-references: EMBL:M30398; NID:G159876; PIDN:AAA29413.1; PID:G159877
C:Genetics:
A:Gene: ovt1
C:Keywords: leucine zipper

Query Match 7.1%; Score 106.5; DB 2; Length 2022;
Best Local Similarity 22.3%; Pred. No. 71;
Matches 66; Conservative 51; Mismatches 106; Indels 73; Gaps 13;

QY 32 DQVIFWKFDETIKLSRFKQYSQASVLDGDKVILLMDSQKYFEATQTVYECGVVT 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 652 DEIRKQK--DKLAKLADLENKLNNETK-MRGDAKLNQRHLDEIDNFKQINYEIVT 708
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 QLLSAVILLFD---EYNEKASAKDILIRI-----LDDGVNKLNEAQKSLGSSQ--- 139
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 709 -IIRQNDQDFDTQMKTNQAKSSMRNLSLIAKKEIEKUSEMNRLOQDQNDLIGAKQGD 767
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 140 -SFNNASGKLALD---SOLTNDPSEKSSYFQSDVRIRKEAYAG-----AAAGIV 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 768 TELNLLTEKIRKVEFERIKKDNQOELEHRTARDLQETNRNHLHLAKELEEARADIV 827
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 AGPGLIISYIAAGVIEGKLIPELNDR-L-KAVQNFFTSLSVTVQANKDIDA-K----- 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 828 A-----LNDRLAKMDANFKIKLDETIIKSPADHETIKSRSEK 864
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 -----LKLATIAAIGIKETETTTTFYVDYDDMLMLLLKGAAKMWINTCNEYQOR 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 865 SEKIVKHETKIYEINKYRAELE--KLSDKDDL-----EKRIIGLODELNEK 910
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 32
T22976
hypothetical protein F59A2.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22976; T23157
R:Lightning, J.
submitted to the EMBL Data Library, June 1994
A:Reference number: Z19645
A:Accession: T22976
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1133 <WIL>
A:Cross-references: UNIPROT:Q21022; EMBL:Z34801; PIDN:CAA84332.1; GSPDB:GN00021; CESP:F5;
A:Experimental source: clone F59A2
R:Burton, J.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19700
A:Accession: T23157
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1133 <WIL>
A:Cross-references: EMBL:Z66514; PIDN:CAA91344.1; GSPDB:GN00021; CESP:F59A2.6
A:Experimental source: clone K01A11
C:Genetics:
A:Gene: CESP:F59A2.6
A:Map position: 3
A:Introns: 13/2, 43/3, 107/3, 413/3, 492/3, 567/3, 635/3, 710/3, 738/3, 795/2, 1008/3; 1
A:Query Match 7.0%; Score 106; DB 2; Length 1133;
Best Local Similarity 18.5%; Pred. No. 36;
Matches 51; Conservative 61; Mismatches 122; Indels 42; Gaps 6;

QY 38 KTFDETIKLSRFKQYSQASVL---VGDIKVLLMDSQKYFEATQTVYECGVVTQ-- 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 KEKEMVQOQLSRLOQOEMLKCDALQAEVNEAKALREEIQAKYDDVTQKAERTQGELESK 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 93 -LLSAVILLFDEYNEKKAQKDIILIRILDDGVNKLNEAQKSLGSSQSFNNASGKLAL 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 KVLSEKQAFENEKEQEREQQLAKAMEKLNSEQNILDEVTKKLEQSEEEVLAAAGAIQEL 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 DSOLTNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPEL 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 TEKL-ESEKETSTAKTELEAVSKLDSSETS-----LKEF 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 212 NDLKAVQNFFTSLSVTVQANKDIDA-KLKLATIAAIGIKETE-----TETTFY 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 SDMIEMK---IQLINCEKQKDEAVELLKQKLEVEKNMSDVEVQKQLLLETTSEMQRH 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 VDYDDLMLLLKGAAKMWINTCNEYQORHGGKKTLL 299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 AEMAEIVKQLEEAQSSIENLKDAENERNLKTAL 284
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 33
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C:Accession: JC5837
```



R.Toki, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.  
Cell Struct. Funct. 22, 565-577, 1997  
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec  
A:Reference number: JC5837; MUID:98093490; PMID:9431462  
A:Accession: JC5837  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-3187 <TOK>  
A:Cross-references: UNIPROT:Q63714; DDBJ:D25543; NID:G516825; PIDN:BAA05026.1; PID:G5168  
C:Comment: This protein plays a role in the formation and maintenance of the characteris  
C:Superfamily: giantin  
F:49-549, 624-1176, 1238-1707, 1763-3114/Domain: coiled-coil leucine zipper #status predict  
F:3165-3187/Domain: membrane anchor #status predicted <MAD>  
  
Query Match 7.0%; Score 106; DB 2; Length 3187;  
Best Local Similarity 22.2%; Pred. No. 1.4e+02;  
Matches 78; Conservative 53; Mismatches 134; Indels 86; Gaps 16;  
  
QY 6 AEQTV--EVVKSIAETADGALDFYNYKLDQVTPWKTFTETTKELSRFKQEQYSQASVLVG 63  
Db 1748 AEATLLANGAKPGVSTPSSHDDINNYLQDLQK---GRIAELEMEKQKDELSTLTLEN 1804  
  
QY 64 DIKVL-----MDSQKYFE-----ATQTVYEWGVTOLLAYILLFDEYNEKKAS 110  
Db 1805 EKNAITQISAKDSEKLEEEVAKINMLNQIQEELSRVTKL-----KETAE 1852  
  
QY 111 AQKDILIRILDGWNKLENAQKSLGSSQSFNNASGKLALDSQLTN-----DPSEKSS 164  
Db 1853 EEKDDLEERL---MNQALENGISGNYQDVTDQAIKNEQLESEMQLKRCVSELEEBKQ 1909  
  
QY 165 YF---QSQVD-RIRKAYAGAAAGVAGPFGILISVSIAAGVIEGKLPINLDRKLVQV- 219  
Db 1910 QLVKEKTKVSEIRKE-YMEKIQGAQPGSKTHAKELQ-----ELLKEQOEVKQLQK 1962  
  
QY 220 -----NFFTSLVTVK-----QANKOIDAKKLATEI-----AAIGEIK----- 254  
Db 1963 DCTRYIGRISALEKTVKALEFVTEFSQKLDATKGNLAQAVEHHKKAQELSSFKILLDD 2022  
  
QY 255 TETETTFRYVD-----YDDLMLSLKGAKKMINTCNEYQQRHGKK 295  
Db 2023 TQSEARVLADNLKKELOSNKESQIKQKDEDLRLRLEQAEKHKE 2073  
  
RESULT 34  
T08164  
dynein alpha heavy chain - Chlamydomonas reinhardtii (fragment)  
C:Species: Chlamydomonas reinhardtii  
C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 02-Feb-2001  
C:Accession: T08164  
R:Mitchell, D.R.; Brown, K.S.  
J. Cell Sci. 107, 635-644, 1994  
A:Title: Sequence analysis of the Chlamydomonas alpha and beta dynein heavy chain genes.  
A:Reference number: Z16302; MUID:94274778; PMID:8006077  
A:Accession: T08164  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2405 <MIT>  
A:Cross-references: EMBL:I26049; NID:9415679; PIDN:AAA57316.1; PID:G603079  
A:Experimental source: strain 21gr  
C:Genetics:  
A:Gene: OD411  
A:Note: Intron positions not resolved (incomplete sequence)  
C:Superfamily: dynein heavy chain, ciliary  
C:Keywords: nucleotide binding; P-loop  
F:575-582/Region: nucleotide-binding motif A (P-loop)  
  
Query Match 7.0%; Score 105; DB 2; Length 2405;  
Best Local Similarity 21.0%; Pred. No. 1.1e+02;  
Matches 75; Conservative 44; Mismatches 110; Indels 128; Gaps 14;  
  
QY 38 KTFDETI---KELSRFKQYSQ-----EASVLVGDIKVLLMDSQDKY 76  
Db 1793 KTFLELIKLVKNVLAARAKRANQNTERLENGHLKHLKQVADVILVEAKVKAVEHVKV 1852

QY 77 FE-----ATQTVYEWGVTOLLAYILLFDEYNEKKASAKQDIL 116  
Db 1853 ASANIPFAEQGVKEKVNENAAQVFAEKCAVIK-----EYSEKQASCEKDLA 1902  
  
QY 117 IR--ILDDGVNKLNEAQKSLGSSQSF-----NNASGKLALDSQL 155  
Db 1903 AAEPLVAENAALETVTWKDLGEAKSLKKPPGVDITAVVILLNNPKKSWOAAQKL 1962  
  
QY 156 TND---FSKSSYFQSDVR-----IRKEAY---AGAAAGTVAGPF 190  
Db 1963 MNVDFLERVSKFSVIDAGQVARKTVDACRPLYALEWFNEALCKKSAAGLCEWAV 2022  
  
QY 191 GLIISYSIAAGV-----IEGKLIPELNDRKLVQNFSTLSVTV-----KOAKND 235  
Db 2023 NIITYDDVQVEPKQKQELAAANAKLEEVNTLAAVEKVALLNKAVQEQYKEANDD 2082  
  
QY 236 IDAA-----KLKLATEIAAIGIKETETETTFRYVD-----YDDLMLSLKGA 278  
Db 2083 KEAARESERCQKLELANRL--INALASEGERWALTVEQLRKSVEVLTGDMLLAA 2137  
  
RESULT 35  
T41342  
probable coiled-coil protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T41342  
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.  
submitted to the EMBL Data Library, January 1999  
A:Reference number: Z21970  
A:Accession: T41342  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1115 <LYN>  
A:Cross-references: UNIPROT:O94488; EMBL:AL035076; PIDN:CAA22653.1; GSPDB:GN00066; SPDB:S  
A:Experimental source: strain 972h-; cosmid c417  
C:Genetics:  
A:Gene: SPDB:SPCC417.07c  
A:Map position: 1  
  
Query Match 6.9%; Score 104.5; DB 2; Length 1115;  
Best Local Similarity 20.6%; Pred. No. 44;  
Matches 66; Conservative 60; Mismatches 127; Indels 67; Gaps 14;  
  
QY 7 EQTVVWKSJAEIETADGALDFYNYKLDQVTPWKTFTETTKELSRFKQEQYSQASVLVGDIL 66  
Db 500 EQSIELTRLYKQLQDIEDYENKLMRMQQWR---EDVDQLQEVVEITQE---LQDTK 552  
  
QY 67 VLLMDS---QDKYFENTQTVYEWGVTOLLAYILLFDE---YNEKKASAKQDILIRIL 120  
Db 553 EYLSKSKESDDYEEV-----GKLTEAEERIEKEFEKTIRENEESISLFEKEVEKLT 605  
  
QY 121 D-----DGVNKLNEAQKSLGSSQSFNNASGKLALDSQLTNDFSEKSSYFQSQV 170  
Db 606 DEITQLSERYNDKCHEPDELQKRL---QTLSEENNAKEDSTKSTNLLQKRWTEAEV 661  
  
QY 171 DRIRKEAVAGAAGIVAGPFGILISYSIAAGVIEGKL-----IPELNDLKAVQNFST- 224  
Db 662 DSLRKENENKOV-----IALKESLVSNSNDKLLNLEQIESLNDQLQKTEMESV 713  
  
QY 225 -----LSVTVKQANKOIDAAKLKAT---EIAAIGEIKTE-TETTFRYV---DY 266  
Db 714 TTSKESLADYLSNLKXERHNDELDSLKKLREPEGILSSNLSLKSEIERNNQYVTLRENF 773  
  
QY 267 DDLMLSLKGAKKMINTCN 286  
Db 774 DSLQNAIME-TFDKQVTHCS 792  
  
RESULT 36  
D84727  
probable RAD50 DNA repair protein [imported] - Arabidopsis thaliana



C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: D84727  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: D84727  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1292 <STO>  
A:CROSS-references: UNIPROT:Q9SL02; GB:AE002093; NID:G4263721; PIDN:AAD15407.1; GSPDB:GN  
C:Genetics:  
A:Gene: At2g31970  
A:Map position: 2  
C:Superfamily: RAD50 protein

Query Match 6.9%; Score 104.5; DB 2; Length 1292;  
Best Local Similarity 15.0%; Pred. No. 53;  
Matches 64; Conservative 93; Mismatches 132; Indels 137; Gaps 13;  
Qy 7 EQTEVVKSAETADGALDFNFKYLDQV-----IPWKTDETIKE 46  
Db ETSQVQVDAEVRHNEKEMMLKDLRLQDQVSIKTAERSTLFEQQRQYAAALPEEN-EDTIEE 304  
Qy 47 LSRFKOYSQBASVLVGDIVL---LMSODKYFEATQTVYEWCGVVTOL---LSAYILL 100  
Db LKWKSKFEERLALLGCTKIRKMEREMVDTETIISLHNAKNTYMLEISKLTQTEAEMML 364  
Qy 101 FDEYNEKASAKQ-----DILIRLDGQVKNLEAKQSLGSSQS--- 140  
Db ---KNERDSTQNTFFHYNLGNVPSTPSTEWLNLNTRIKSLGCELEMDLLDKKSNET 421  
Qy 141 -----FNASGKLLALDSQ-----LTNDSEKSSY----- 165  
Db 422 ALSTAWDCYMDANDRWKISIAQKRAKBIKMGISKRIBERDSEFEISTVDVKQTD 481  
Qy 166 -----FOSQVDRIKEAVA-----GAAAGIVAGPFGLIISY 196  
Db 482 EREKQVQVELRKTKQNSERGFESKIEQKQHEIYSLHKKITLNRDVRVAGDREDRLLT 541  
Qy 197 SI-----AAGVIEGKLIP-----ELNDRKAVQNFTSLSVTVKQAKNDIDAAKKLA 244  
Db 542 RIDECKDIRGVKGLPPEKDMKREIVQALRSIEREYDDLRLSKSREAEKVNMLQMKIQ 601  
Qy 245 TEIAAIGIKETETTRFYVDYD-----DLMLSLKGAQAKMINTCNEYQQRHG 293  
Db 602 EVNNSLFKHNKDTESKRYIESKLQALKQESVTIDAYPKLJESAKDKDRDRKREYNMANG 661  
Qy 294 KKTLL 299  
Db 662 MRQMF 667

RESULT 37  
A59282  
nonmuscle myosin II heavy chain A - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jul-2004  
R:Accession: A59282  
R:Bhati, Dev, N.; Taira, M.; Nooruddin, H.; Adelstein, R.S.  
Mech. Dev. 78, 33-36, 1998  
A:Title: Differential expression of non-muscle myosin heavy chain genes during Xenopus e  
A:Reference number: A59282; MUID:99077683; PMID:9858676  
A:Accession: A59282  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1964 <BHA>  
A:CROSS-references: UNIPROT:Q93522; GB:AF055895; NID:G3660671; PIDN:AAC03556.1; PID:G366  
A:Experimental source: cell line XTC  
C:Superfamily: myosin heavy chain; myosin motor domain homology

F:84-764/Domain: myosin motor domain homology <MMO>  
Query Match 6.9%; Score 104.5; DB 2; Length 1964;  
Best Local Similarity 19.8%; Pred. No. 93;  
Matches 63; Conservative 47; Mismatches 103; Indels 105; Gaps 13;  
Qy 7 EQTEVVKSAE-----TADGALDFNFKYLDQVIPWKTDETIKE-----LSRFKQYS 55  
Db 1166 EQETHLTKLTLEDEARHGEIQIRQKHSQAV--EELSELEQTKRLKGNLEKAKQALE 1223  
Qy 56 QEASVLVGDIVKLLMSODKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASQ-KD 114  
Db 1224 GERELANEVTLQKGD-----SEHKKKVEAQLQE 1256  
Qy 115 ILIRI-----LDDGVNKNL---EAKSLLGSSQSFNNASGKLLALDSQLTDFSEK 162  
Db 1257 LQVKVTEGDRVRSSELSEKANRLQVELDNVNSLSQSDSKSIKLGK-----DFSTL 1306  
Qy 163 SSYFQSQVDRIKRAYAGAAAGIVAGPFGLIISYIAAGVIEGK---LIPELNDRLKAVQ 219  
Db 1307 ESQFQDAQELLQEETROK-----LSFSTKLQMEDEKNGLLLEQLEEEAAK 1353  
Qy 220 NFPTSLV-----TV-----KOANKDIDAALKLATAIAIGETKT 255  
Db 1354 NLCKQISTLSQSMQMDMKKMDENVGSLETVEELKKLQKLEAVNQRFEEKAAAYD--KL 1411  
Qy 256 ETETTFYVDYDDLMLSL 273  
Db 1412 EKTTRLQQLDLDISVDL 1429  
RESULT 38  
T03754  
hypothetical protein E - slime mold (Physarum polycephalum) mitochondrion integrated plas  
C:Species: mitochondrion Physarum polycephalum  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T03754  
R:Nakagawa, C.C.; Jones, E.P.; Miller, D.L.  
Curr. Genet. 33, 178-187, 1998  
A:Title: Mitochondrial DNA rearrangements associated with mf plasmid integration and pla  
A:Reference number: 215055; MUID:98177147; PMID:9508792  
A:Accession: T03754  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-640 <NAK>  
A:CROSS-references: UNIPROT:Q35593; EMBL:AF012249; NID:G2435419; PIDN:AAC15934.1; PID:G24  
C:Genetics:  
A:Genome: mitochondrion integrated plasmid mf  
C:Keywords: mitochondrion

Query Match 6.9%; Score 104; DB 2; Length 640;  
Best Local Similarity 22.5%; Pred. No. 23;  
Matches 61; Conservative 40; Mismatches 104; Indels 66; Gaps 11;  
Qy 10 VEVVKSATETADGALDFNFKYLDQVIPWKTDETIKELSRFKQYSQBASVLVGDIVKLL 69  
Db 182 LDIMSASRTAD---QLYNNYIQLVNLNLTPLTT-----GVEKVT 219  
Qy 70 MDSQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASAKQDILIRILDG----- 123  
Db 220 KSS-----LSEKVKMLGYA--LLAGAVAFIGWTKYQNVNVLKADGSEIVDNGKRTQV 271  
Qy 124 VNKLNEAKQSLGSSQSFNNASGKLLALDSQ-----LTNDFSEKSSYFQSQVDRI 173  
Db 272 VHGVSETDTLLSATNKAVGVAEKLDDVSTKVDTLGVKHAATLTVNNNTVLFQTLDRV 331  
Qy 174 RKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVKQAN 233  
Db 332 R--GVLELQIQNDAEPLP-----QQAAILQGVNPEL---LGAINHQIESLAVGLKTN 379  
Qy 234 KDIDAANKLATEIA-----AIGEIKTETTT 260  
Db 380 NDLS-----KLANSILNLENQIKETIAAQRVT 406

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, A.; Tanaka, M.; Tamakoshi, A.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. *Bacillus subtilis*.  
A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: E69730  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-2334 <KUN>  
A;Cross-references: GB:Z99124; GB:AL009126; MID:g2636442; PIDN:CAB15959.1; PID:g2636469  
A;Experimental source: strain 168  
R;Toshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.  
Microbiology 141, 337-343, 1995  
A>Title: Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome contain  
A;Reference number: Z24350; MUID:95219088; PMID:7704263  
A;Accession: T47101  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Residues: 1-2334 <YOS>  
A;Cross-references: EMBL:D31856; NID:G603765; PIDN:BAA06656.1; PID:G603782  
A;Experimental source: strain BGSC1A1  
C;Genetics:  
A;Gene: wapa; N17G  
C;Superfamily: cell wall-associated protein wapa  
Query Match 6.9%; Score 104; DB 2; Length 2334;  
Best Local Similarity 25.6%; Pred. No. 1.3e+02;  
Matches 60; Conservative 39; Mismatches 85; Indels 50; Gaps 13;  
Qy 78 EATQTV-YEWGVVTVLLSAYILLFDEYNEKK-ASAQKILIRILD---DGVNKLNEAOK 132  
Db 1203 KATESYQYDQGNVTGKAYGTETTYEYNNKNDVTGKQTEGNTVDIAVDGLDVSSETDQ 1262  
Qy 133 SLIGSS-----QSFNN--ASGKLLALDSQLTNDPS-----EKSYY-FQSQVDRIKREAYAGAA 182  
Db 1263 SGKSSAAVYDYGNGQIQSSKDLASTNLIKQSFQAQSGWNLTKSKDR-RKISVIADK 1321  
Qy 183 AGIVAGPFGI-IISYSIAAGVIEGKILPELNDRLKAVQNFFTSLSVTVKQAKNDIAAKL 241  
Db 1322 SGVLGSKALEVLVSQTSAGTDHG-----YSSATQTV-----EL 1355  
Qy 242 KLATEIAAIGIKETETTFYDYDDMLSLKLLGAAKMMKNTC-NEYQORHGK 294  
Db 1356 EPNTYTLGSKIKTDLAKRAYFNID-----LRDQDKQKIQWHNEYSALAGK 1403  
RESULT 41  
G97236  
A;Parse involved in DNA repair [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: G97236  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: G97236  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1163 <KUR>  
A;Cross-references: UNIPROT:Q97FK1; GB:AE001437; PIDN:AAK80682.1; PID:g15025772; GSPDB:G15025772; GSPDB:G15025772  
A;Experimental source: Clostridium acetobutylicum ATCC924  
C;Genetics:  
A;Gene: CAC2736  
Query Match 6.9%; Score 103.5; DB 2; Length 1163;  
Best Local Similarity 21.7%; Pred. No. 54;  
Matches 81; Conservative 55; Mismatches 142; Indels 95; Gaps 14;

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, A.; Tanaka, M.; Tamakoshi, A.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. *Bacillus subtilis*.  
A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: E69730  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-2334 <KUN>  
A;Cross-references: GB:Z99124; GB:AL009126; MID:g2636442; PIDN:CAB15959.1; PID:g2636469  
A;Experimental source: strain 168  
R;Toshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.  
Microbiology 141, 337-343, 1995  
A>Title: Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome contain  
A;Reference number: Z24350; MUID:95219088; PMID:7704263  
A;Accession: T47101  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Residues: 1-2334 <YOS>  
A;Cross-references: EMBL:D31856; NID:G603765; PIDN:BAA06656.1; PID:G603782  
A;Experimental source: strain BGSC1A1  
C;Genetics:  
A;Gene: wapa; N17G  
C;Superfamily: cell wall-associated protein wapa  
Query Match 6.9%; Score 104; DB 2; Length 1496;  
Best Local Similarity 19.0%; Pred. No. 70;  
Matches 62; Conservative 67; Mismatches 124; Indels 74; Gaps 13;  
Qy 4 IFAEQTVVYVKSATETADGA--LDFNKKYLDQVWPWT-----FDETIKLSRP-----KQ 52  
Db 153 IEAEKTVKGMKEMGRDVRVWMEBEKSQVEEKLKWKKEQFKHLEAYEKLKNUFKDSKK 212  
Qy 53 EYSQBSALVGDIKVL--LMSQDKYFEATQTVYEMC-GVVTQLLSAYILLFDEYNEKKA 109  
Db 213 EWEBEKSLDEIYSIQTLKDSVTRISEDLQKLLQNCALQ-----EETRRKHL 263  
Qy 110 SAQKDLIRILDGVNKLNEAOKSLGSSQSFNAGKLLALDSQLTNDPSFKSYFO-- 167  
Db 264 EIQVSEFKAYEDAFACQDARTOL-----DDLAKGKDWEVAELRQTLMSKMDAYFKEM 316  
Qy 168 -----SQVDRTKAYAGAAAGIVAGPFGILLISVIAAGVIEG---KLIPELNDRLKAVON 220  
Db 317 KYENGKLEQENRELLGSLK-----ELQENIQSGNSALSGLKKNKFNLEN 362  
Qy 221 FFTSLSVT-----VKQANKDIDAAKLKLAIEAIGIKETETE-----TTRFTV 264  
Db 363 IHKNCANLSKAEWSSQVEKVEEINDYKQLQSKAEALKEVELENCRSSTAQML 422  
Qy 265 DYDDLMLSLKGAAKMMKNTCNEYQOR 291  
Db 423 QYEBISIMFL-----VLSRTVSEAQR 444  
RESULT 40  
S32920  
cell wall-associated protein precursor wapa [similarity] - Bacillus subtilis  
C;Species: Bacillus subtilis  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C;Accession: S32920; E69730; T47101  
R;Foster, S.J.  
Mol. Microbiol. 8, 299-310, 1993  
A>Title: Molecular analysis of three major wall-associated proteins of *Bacillus subtilis* protein.  
A;Reference number: S32919; MUID:93302506; PMID:8316082  
A;Accession: S32920  
A;Molecule type: DNA  
A;Residues: 1-2334 <FOS>  
A;Cross-references: UNIPROT:Q07833; GB:L05634; NID:g304177; PIDN:AAA22883.1; PID:g304179  
R;Kunat, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet, C.; Bron, S.; Brouillet, S.; Bruchet, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997  
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Konningstein, G.; Krogh, H.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel

QY 1 MTGIFAEOTVVEVKSIAETADG-----ALDFYNYKLDQVTPWKTDETIK-----ELS 48  
Db 112 ITGDEVELEEGAKSVNEKQOEIIIGLSLDDFTRTV--VLPQCKFSEFLKRGKERRNMLE 169  
QY 49 RF--KQEYSQAS-----VLGDIKVLMLDSQDKYFEATQTVYEWCGVVT 91  
Db 170 RLFNLOEYGDLSFKLARKIRKREKENVLGELGYENINEDVLKERRELLKENNDFN 229  
QY 92 QLLSAYIILPDYNEKKAQAOKDILI-----RILDDGVNKLNE-----AOKSLIGSSQS--- 140  
Db 230 EASKEYLKABEYNEGVEKVGLOIEIEBKNRVRKOLMEKKDEIDLKERRARLGSSSKVK 289  
QY 141 -----FNWASKLLALDSQLTN-----DFSEKSSYFQSDVRIRKAYAGAA 182  
Db 290 PYIDNYENTLQIDILKEQILSRENTMKAISLEKEDMEKLSIAKDNNKEK----- 339  
QY 183 AGIVAGPFGLIISYSIAAGVLEGKLIPELNDRLKAVQNFPTSLSV-----TVKQANKD 235  
Db 340 -----ALPFWMKHIIIDAIKEKDLDDNIKLEKKRLOQKIEKLSLEASNKBELIKQNIKD 395  
QY 236 IDAAKLKLATIAAIGEIKTETTRTFYVDYDMLSLKGAKKMIN-----TCNEYQ 289  
Db 396 IDSLTLKIQNLKESKIDNLKVPKE-----YKNKINEGIFLLRNYDEKLSKHKHKLGLDCKDFQ 451  
QY 290 -----QRHGCKTLL 298  
Db 452 VDFEKAQSKKEML 464

RESULT 42  
T28667  
dynein beta heavy chain - Paramesidium tetraurelia  
C:Species: Paramesidium tetraurelia  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T28667; PC4340  
R:Kandl, K.A.; Forney, J.D.; Asai, D.J.  
submitted to the EMBL Data Library, January 1995  
A:Description: The dynein genes of Paramesidium: the differential expression of axonemal a  
A:Reference number: Z20502  
A:Accession: T28667  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-4588 <KAN>  
A:Cross-references: UNIPROT:Q94709; EMBL:U19464; NID:G625089; PID:G625090; PIDN:AAA61680  
R:Asai, D.J.; Beckwith, S.M.; Kandl, K.A.; Keating, H.H.; Tjandra, H.; Forney, J.D.  
J. Cell Sci. 107, 839-847, 1994  
A:Title: The dynein genes of Paramesidium tetraurelia: Sequences adjacent to the catalytic  
A:Reference number: PC4340; MUID:94334383; PMID:8056840  
A:Accession: PC4340  
A:Molecule type: mRNA  
A:Residues: 1886-2085 <ASA>  
C:Genetics:  
A:Genetic code: SGC5  
A:Introns: 43/2; 113/3; 171/3; 4490/1  
C:Superfamily: dynein heavy chain, ciliary  
C:Keywords: nucleotide binding; P-loop  
F:1886-1893/Region: nucleotide-binding motif A (P-loop)

Query Match 6.9%; Score 103.5; DB 2; Length 4588;  
Best Local Similarity 19.3%; Pred. No. 3.3e+02;  
Matches 69; Conservative 67; Mismatches 115; Indels 107; Gaps 15;

QY 7 EQTVVEVKSIAETADGALDFYNYKLDQVTPWKTDETIKELSRFPKQ 52  
Db 1207 QETINIKKKLDEFOKKVSDFRNDTLNNLPYSVHDDMKMDQILYSVTIDEYKLLQNEK 1266  
QY 53 E---YSQASVLVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAVILLFD----- 102  
Db 1267 EAADYNQ-----LEKLEFEKSGYKQLRNETNVDLSKLSKIMWDALSMVNY 1310  
QY 103 EYNEKKA---SAQKDILI---RILDDGVNKLNEAQAOKSLIGSSQSFPNNASGK---LLALDS 153  
Db 1311 QYNDWKSKEFRQIKADVILLESKNVLGNQLKNLPKEVRNPKGYNIAIVDKRNMSVVLPLVS 1370

QY 154 QLTNDFSEKSSYFQSDVRIRKAYAGAAAGIVAGPFGLIISYSIAAGVIRGKLIPELND 213  
Db 1371 ALHSEFMEDRHW--SQVKDMTKSKF-----SHKAMTFLFD 1403  
QY 214 RLKAVQ--NPFTSLSVTVKQANKDIDAOKLKLATIAAIGEI--KTETETTRFYVDYDDL 269  
Db 1404 DILALQLYFDQAQNEVVEVASKEAKTEKLMETATLWKQIFEFEDYKETKVFPLDNNM 1463  
QY 270 MLSL-----LKGAACKM---INTCNEYQQRHGK-----KTLLEV 300  
Db 1464 MEMLDQHSLDLMGMKGOKGYVEFYNTVDWRKGLRVDSVVGWELVKVQKWKTLVNI 1521

RESULT 43  
A36014  
myosin heavy chain, nonmuscle - fruit fly (Drosophila melanogaster)  
N:Contains: myosin ATPase (BC 3.6.4.1)  
C:Species: Drosophila melanogaster  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C:Accession: A36014; B36014  
R:Ketchum, A.S.; Stewart, C.T.; Stewart, M.; Kiehart, D.P.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6316-6320, 1990  
A:Title: Complete sequence of the Drosophila nonmuscle myosin heavy-chain transcript: cor  
A:Reference number: A36014; MUID:90349606; PMID:2117279  
A:Accession: A36014  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-2017 <KET>  
A:Cross-references: UNIPROT:Q99323; GB:M35012  
C:Genetics:  
A:Gene: FlyBase:zip  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated  
F:1-2017/Product: myosin heavy chain, form I #status predicted <MA>  
F:46-2017/Product: myosin heavy chain, form II #status predicted <MA2>  
F:135-815/Domain: myosin motor domain homology <MMOT>  
F:225-232/Region: nucleotide-binding motif A (P-loop)  
F:598-631/Region: actin binding #status predicted  
F:705-727/Region: actin binding #status predicted  
F:888-2017/Domain: coiled coil #status predicted <COI>  
F:888-1328/Region: S2  
F:1329-2017/Region: light meromyosin  
F:176/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
F:231/Binding site: ATP (Lys) #status predicted  
F:745,755/Active site: Cys #status predicted

Query Match 6.8%; Score 103; DB 1; Length 2017;  
Best Local Similarity 23.8%; Pred. No. 1.2e+02;  
Matches 65; Conservative 44; Mismatches 102; Indels 62; Gaps 13;

QY 8 QTVEVKSIAETADGALDFYNYKLDQVTPWKTDETIKELS-RFKQEYSQ--EASVLVD 64  
Db 1357 ENLRKAKTVLEKAGTLEAEN--ADLATELRSVNSSQENDRRRRKQAESQIAELQVKLAE 1314  
QY 65 IKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKAQAOKDILIRILDGV 124  
Db 1315 IERARSELQEK---CTKLQEAENITNL-----EEAELKASA---AVKSASNME 1358  
QY 125 NKLNEAOKSL-----LGSSQSPNNASGKLLALDSQLTNDFSEKSSY-----FQSQVD 171  
Db 1359 SQLTEAQOLLEETFRQKLGSLSKLRQIESEKAEQLEEDDEAKRNYERKLAEVTTOMQ 1418  
QY 172 RIRKEAYAGAAAGIVAGPFGLIISYSIAAGVIRGKLIPELNDRLKAVQNFPTSLSVTKQ 231  
Db 1419 EIRKKAEDA-----DLAKELEEK--KRLNWDIEALER-----QVKE 1454  
QY 232 ---ANKDIDAOKLKLATIAAIGEIKTETTR 261  
Db 1455 LIAQNDRLDKSKKIQSELED-ATIELEAQRTK 1486



Db 1599 EOSATFNNAVER-----AAQFGAEDALNASATRFSESQAADMVSSSRLEEG 1650  
Qy 206 KLIPBLNDRK-----AVQFFTSLSVTVKQAKDIDAAKLKLATEIAAIGIK 254  
Db 1651 KI-----DRLSNISQTLAQVAGIVGRBEHSKVLQSASELLNAAQSSL-----VGTLE 1699  
Qy 255 TETETTRYVDYDMLSLKGAANK-----MINTONEVOOR 291  
Db 1700 -----BQDALRSLSVLGRKRSBEIETAMRVVVVNTLNEAER 1740  
RESULT 47  
D97198  
methyl-accepting chemotaxis protein [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: D97198  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: D97198  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-379 <KUR>  
A;Cross-references: UNIPROT:Q97GE7; GB:AE001437; PIDN:AAK80375.1; PID:gl5025436; GSPDB:C  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC2420

Query Match 6.8%; Score 102.5; DB 2; Length 379;  
Best Local Similarity 21.2%; Pred. No. 14;  
Matches 58; Conservative 40; Mismatches 83; Indels 93; Gaps 13;  
Qy 72 SODKYFEATQTYEWCVVTTQLLSAYILLFDEYNEK-----KASAKQDI-- 115  
Db 51 SNDEFGNLTHSFF-----TMSQNLKLLIIQIKESSRLEDVSTEMLKSCSESASASSDISQ 106  
Qy 116 -LIRLDD--GVNKLNEAKSLIG-----SSQSFNNAAGK----- 147  
Db 107 NIVNVADSNVSLNKLKANDSVLMLIDVFSQVILNVNIIICKVEETSYVSKGKQNDINK 166  
Qy 148 ----LLALDSQTLNDFSEKSSYFQSOVDRIKAYAGAAAGIV-----A 187  
Db 167 ITEQMLNIENRV-NSLSEVISKLTCHDINN--YSNWISDIAETNLLSNANIEAARA 223  
Qy 188 GPFGLLIISYTAAGVIE-----GKLIPELN-DRLKAVQNF-----FTSLSV 227  
Db 224 GENGK--GFSVVADKIRSLADESGKAASEIGNLIREMNEISIKAINSMNAGKKEVTSGKV 281  
Qy 238 TVQKQAKDIDAAKLKLATEIAAIGIKTETETTR 261  
Db 282 IVEQNLTFNEIANSINSILAATSEVTASMTVK 315

RESULT 48  
H81665  
replicative DNA helicase TC0784 [imported] - Chlamydia muridarum (strain Nigg)  
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C;Accession: H81665  
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A;Reference number: A81500; MUID:20150255; PMID:10684935  
A;Accession: H81665  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-472 <TET>  
A;Cross-references: UNIPROT:Q9PJP4; GB:AE002346; GB:AE002160; NID:g7190805; PIDN:AAF3958  
A;Experimental source: strain Nigg (MoPn)

C;Genetics:  
A;Gene: TC0784  
C;Superfamily: phage P22 gene 12 protein  
Query Match 6.8%; Score 102.5; DB 2; Length 472;  
Best Local Similarity 23.1%; Pred. No. 19;  
Matches 74; Conservative 42; Mismatches 105; Indels 99; Gaps 17;  
Qy 25 DFVNYKLDQVIPWKTFFDETIK-----ELSRFKQESYQSEASVLLVGDIDKVLMDMS 72  
Db 48 DFY--FLEHRIIFRVLDQAFKSDRPMPLHTGEELKR-----RDQLNVIGASVLI--- 96  
Qy 73 QDKYFEATQTYEWCVVTTQLLSAYILLFDDYNE-----KASAKQDILIRILD-- 121  
Db 97 -----TLSEFAGT-----SAYI---EEYADIIRSKSILRKMIQAAKDIEKKAEEP 139  
Qy 122 -DGVNKLNEAKSKLLGSSQSFNNA-----SGKLALDSQTLNDF-----SEKSSYFOSQVD 171  
Db 140 RDVTALDDAQNLLFRISQITNFAPYVLSVDKLGSLSTKDKSFLALQERQEAFOASSH 199  
Qy 172 RIRKEAYAGAAAGIV-----AGPFGLLIISYTAAGVIEGKLIPELNDRLKAVQNF-- 221  
Db 200 DARIPTLSGGPTHTFLDLDRMLNGFSPNLI---LAARPAMGKTALALN-----IVENFCF 252  
Qy 222 -----FTSLSVTVKQ-----ANKDIDAAKLKLATEIA-----AIGIKTETETTR 261  
Db 253 ESRLPVGITFSLMTVTDQLIHRICSRSEVEAKISVG-DISGRDFQRVVSVVREMEHTL 311  
Qy 262 FYVDYDDLMLSLKGAAKKM 281  
Db 312 LIDDYPLGKITDLRARARM 331

RESULT 49  
T44825  
hypochemical protein wzc [imported] - Acinetobacter lwoffii  
N;Alternate names: protein tyrosine kinase  
C;Species: Acinetobacter lwoffii  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T44825  
R;Nakar, D.; Gutnick, D.L.  
submitted to the EMBL Data Library, July 1999  
A;Description: Genomic organization of the wzc region of Acinetobacter lwoffii RAG-1 req  
A;Reference number: Z22856  
A;Accession: T44825  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-726 <NAK>  
A;Cross-references: UNIPROT:Q9RMD9; EMBL:AJ243431; PIDN:CAB57193.1  
A;Experimental source: strain RAG-1  
C;Genetics:  
A;Gene: wzc

Query Match 6.8%; Score 102; DB 2; Length 726;  
Best Local Similarity 21.7%; Pred. No. 36;  
Matches 67; Conservative 48; Mismatches 92; Indels 102; Gaps 13;  
Qy 8 QTVBWKSAIETADGALDFVNYKLDQVIPWKT--FDETIKLSRFKQESYQSEASV----- 60  
Db 173 KTDEIVLTA-----PLNKLNLQKNQGTWKVAFITNDTFDANVYFVKNSLPAAVQAIL 226  
Qy 61 -----LVGDIKVLMDSDQKYFEATQTYEWCVVTTQLLSAYILLFDEYNEKKASA 111  
Db 227 NYSVAERGKLTG-ILGLNYQODKEH-----ITKVNAILAAYSAQNVERRSA 273  
Qy 112 QKQDILIRILDGV-----NKLNEAKSKLLGSSQSFNNAAGKLLALDSQTLNDFSEKSSYFQ 167  
Db 274 ETAQTLFLDEQLDPLKQLDDAERVFNFKQQYN-----TVDVTKESELYL 320  
Qy 168 SOVDRIR-----KEAYAGAAAGIVAGPFGLLIISYTAAGVIEGKLIPELNDRLKAVQNF 221  
Db 321 TQSTLTETKIELOQQOQADLAUKYTA-----EHPAIREINAQITALDKQ 364

Wed Feb 2 11:26:41 2005

QY 222 FTSLSVTVKQANKDIDAALKLATEIAAIGAIKIKETETETTFYVDYDDLMLSLKGAKKM 281  
Db 365 ITQLNSTLUQL-PDIQRQYLQFLREV-----EVKTQ-----LYTALL----- 400

QY 282 INTCNEYQQ 290  
Db 401 ----NSYQQ 405

RESULT 50

H75001  
methyl-accepting chemotaxis protein (tlpc) PAB1336 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C:Accession: H75001  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru  
A:Reference number: A75001  
A:Accession: H75001  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-739 <RAW>  
A:Cross-references: UNIPROT:Q9UYF8; GB:AJ248288; GB:AL096836; NID:95458960; PIDN:CAB5045  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB1336

Query Match 6.8%; Score 102; DB 2; Length 739;  
Best Local Similarity 19.5%; Pred. No. 37;  
Matches 64; Conservative 73; Mismatches 140; Indels 52; Gaps 11;  
QY 1 MTGIFAEQTVVVKSA--IETADGALDFYKVKLDQVIPKWTDFETIKELSRFKQEYSQEA 58  
Db 425 VTETFERISGLVEMANDLEKRANALQVSKDVTEAI--NQVNEAIQQVSIQAQOQETI 482  
QY 59 SVLVGDIKVLMDSQDKYFEATQTVVCGVVTOLLISAVILLDFEYNEKKASAKDI--- 115  
Db 483 NEITDGMRLVAQTS-----ESVRAMEEFGSNAVTEVVS-----IANEGSQKGDALAKRIEDI 534  
QY 116 --LIRILDGWNKLINE-----AQKSLGSSQSFNNA-----SGKLLALDS 153  
Db 535 QHMSRIEETVSKVAEMSRNIEITNVITSIAEQTNLLALNAAIEAARAGEAGRGFAVVA 594  
QY 154 QLTNDFSEKSSY-----FQSQVDIR--KEAYAGAAGI-VAGPFGLIISYSIAAGVIEG 205  
Db 595 QEIRKLAEESKQADNIKSIIIDKIDKEAVEATKEGVSVIGESSEITLRTDITIGYLANIA 654  
QY 206 KLIPELNDRLKAVQNFFTSLSVTVKQANKDIDAALKLATEIAAIGAIKIKETETTFYVD 265  
Db 655 TLLQETSERM-----TTVKEQIVRTQEEDVKALRALENLAASABETTASAEVSSAIE 707  
QY 266 YDDLMLSLKGAKKMINTCNEYQORHGK 294  
Db 708 QOTAAIEELRRAAQELKMDVGRMEQIVGK 736

Search completed: January 28, 2005, 19:16:48  
Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:06:39 ; Search time 198 Seconds  
(without alignments)  
880.498 Million cell updates/sec

Title: US-09-993-292B-24  
Perfect score: 1510  
Sequence: 1 MTGIFAEQTVVKSIAETA.....TCNEYQQRHGKTLLEVPDI 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : UniProt 02.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID         | Description          |
|------------|-------|---------------|--------|----|------------|----------------------|
| 1          | 1505  | 99.7          | 302    | 1  | HLVE_SALPA | Q93rr6 salmonella    |
| 2          | 1467  | 97.2          | 302    | 1  | HLVE_SALTI | Q8z727 salmonella    |
| 3          | 1376  | 91.1          | 302    | 1  | HLVE_ECOLI | P77335 escherichia   |
| 4          | 1361  | 90.1          | 302    | 1  | HLVE_ECO57 | Q9reb3 escherichia   |
| 5          | 1104  | 73.1          | 300    | 2  | Q9X258     | Q9x268 escherichia   |
| 6          | 513   | 34.0          | 113    | 1  | HLBL_SHIFL | Q9rct3 shigella fl   |
| 7          | 382   | 25.3          | 93     | 1  | HLBL_ECOL6 | Q8fi27 escherichia   |
| 8          | 136.5 | 9.0           | 895    | 2  | Q9LIW7     | Q9liw7 oryza sativ   |
| 9          | 132   | 8.7           | 495    | 2  | P71497     | P71497 mycoplasma    |
| 10         | 130   | 8.6           | 2723   | 2  | Q7ROB6     | Q7rgb6 mycoplasma    |
| 11         | 128.5 | 8.5           | 713    | 2  | Q6MT03     | Q6mt03 mycoplasma    |
| 12         | 128.5 | 8.5           | 713    | 2  | CAE77235   | CAE77235 mycoplasma  |
| 13         | 126.5 | 8.4           | 577    | 2  | Q7N5U5     | Q7n5u5 photorhabd    |
| 14         | 125   | 8.3           | 1521   | 2  | Q6G015     | Q6g015 bartonella    |
| 15         | 124.5 | 8.2           | 1825   | 2  | Q7RIT2     | Q7rit2 plasmodium    |
| 16         | 122.5 | 8.1           | 4007   | 2  | Q7SHZ4     | Q7shz4 neurospora    |
| 17         | 121   | 8.0           | 2664   | 2  | Q7REL0     | Q7rel0 plasmodium    |
| 18         | 120.5 | 8.0           | 693    | 2  | Q7N9I8     | Q7n9i8 wolinnella s  |
| 19         | 117.5 | 7.8           | 555    | 2  | Q7VI37     | Q7vi37 helicobacte   |
| 20         | 117   | 7.7           | 2139   | 2  | Q07569     | Q07569 entamoeba h   |
| 21         | 116.5 | 7.7           | 478    | 2  | Q64067     | Q64067 bacterioph    |
| 22         | 116.5 | 7.7           | 478    | 2  | Q31954     | Q31954 bacillus su   |
| 23         | 116.5 | 7.7           | 652    | 2  | Q7CMF0     | Q7cmf0 bacillus an   |
| 24         | 116.5 | 7.7           | 652    | 2  | Q9X360     | Q9x360 bacillus an   |
| 25         | 116.5 | 7.7           | 652    | 2  | AAT28865   | AAT28865 bacillus    |
| 26         | 116.5 | 7.7           | 779    | 2  | Q7RM79     | Q7rm79 plasmodium    |
| 27         | 116   | 7.7           | 541    | 2  | AAR34750   | Aar34750 geobacter s |
| 28         | 116   | 7.7           | 541    | 2  | AAR34750   | Aar34750 geobacter s |
| 29         | 116   | 7.7           | 941    | 2  | Q81Z98     | Q81z98 bacillus an   |
| 30         | 116   | 7.7           | 941    | 2  | AAT29469   | AAT29469 bacillus    |
| 31         | 116   | 7.7           | 1023   | 1  | HLV1_ECOLI | P09983 escherichia   |

|     |       |     |      |   |             |                      |
|-----|-------|-----|------|---|-------------|----------------------|
| 32  | 115   | 7.6 | 1005 | 1 | RA50_METJA  | Q58718 methanococc   |
| 33  | 115   | 7.6 | 1093 | 2 | Q288G7      | Q288g7 listeria in   |
| 34  | 114.5 | 7.6 | 821  | 2 | Q08581      | Q08581 saccharomyc   |
| 35  | 113.5 | 7.5 | 587  | 2 | Q7WH8       | Q7wh8 sulfobolus     |
| 36  | 113.5 | 7.5 | 1478 | 2 | Q7RD20      | Q7rd20 plasmodium    |
| 37  | 112.5 | 7.5 | 520  | 1 | REC_N_AQUAE | Q66834 aquifex aeo   |
| 38  | 112.5 | 7.5 | 584  | 1 | SIR_SINY3   | Q55486 synecocyst    |
| 39  | 112.5 | 7.5 | 941  | 2 | Q6CGC3      | Q6cgc3 yarrowia li   |
| 40  | 112.5 | 7.5 | 1024 | 1 | HLV1_ECOLI  | P08715 escherichia   |
| 41  | 112.5 | 7.5 | 1098 | 2 | Q50733      | Q50733 borrelia bu   |
| 42  | 112.5 | 7.5 | 1906 | 2 | Q6C359      | Q6c359 yarrowia li   |
| 43  | 112.5 | 7.5 | 3102 | 2 | Q45614      | Q45614 caenorhabdi   |
| 44  | 112   | 7.4 | 1012 | 2 | Q6GNE7      | Q6gne7 xenopus lae   |
| 45  | 112   | 7.4 | 1081 | 2 | Q7RFQ7      | Q7rfq7 plasmodium    |
| 46  | 112   | 7.4 | 1464 | 2 | Q81IF6      | Q81if6 plasmodium    |
| 47  | 111   | 7.4 | 329  | 1 | ESG2_TRYBB  | P04478 trypanosoma   |
| 48  | 111   | 7.4 | 329  | 2 | Q6ZM87      | Q6zm87 trypanosoma   |
| 49  | 111   | 7.4 | 329  | 2 | CAD21890    | Cad21890 trypanoso   |
| 50  | 111   | 7.4 | 465  | 2 | Q73AJ3      | Q73aj3 bacillus ce   |
| 51  | 111   | 7.4 | 465  | 2 | AAS40708    | Aas40708 bacillus ce |
| 52  | 111   | 7.4 | 1033 | 2 | Q739A1      | Q739a1 bacillus ce   |
| 53  | 111   | 7.4 | 1033 | 2 | AAS41161    | Aas41161 bacillus    |
| 54  | 111   | 7.4 | 1189 | 2 | Q8CPI0      | Q8cpio staphylococ   |
| 55  | 111   | 7.4 | 1295 | 2 | Q22257      | Q22257 caenorhabdi   |
| 56  | 111   | 7.4 | 2033 | 2 | Q7XEH4      | Q7xeh4 oryza sativ   |
| 57  | 110.5 | 7.3 | 1489 | 2 | Q7RJ32      | Q7rj32 plasmodium    |
| 58  | 110.5 | 7.3 | 2867 | 1 | RBP2_PLAVB  | Q00799 plasmodium    |
| 59  | 110   | 7.3 | 411  | 2 | Q8D3E6      | Q8d3e6 wiggleswort   |
| 60  | 110   | 7.3 | 957  | 2 | Q7RNH1      | Q7rnh1 plasmodium    |
| 61  | 110   | 7.3 | 1285 | 2 | Q9XUJ3      | Q9xuj3 thermotoga    |
| 62  | 109.5 | 7.3 | 504  | 2 | Q8I752      | Q8i752 plasmodium    |
| 63  | 109.5 | 7.3 | 1024 | 2 | Q8G9Z4      | Q8g9z4 escherichia   |
| 64  | 109.5 | 7.3 | 1024 | 2 | Q8GA40      | Q8ga40 escherichia   |
| 65  | 109.5 | 7.3 | 1024 | 2 | Q8FE01      | Q8fe01 escherichia   |
| 66  | 109.5 | 7.3 | 2749 | 2 | Q7REY3      | Q7rey3 plasmodium    |
| 67  | 109   | 7.2 | 541  | 2 | Q7X866      | Q7x866 oryza sativ   |
| 68  | 109   | 7.2 | 541  | 2 | CAE01868    | CAE01868 oryza sat   |
| 69  | 109   | 7.2 | 573  | 2 | Q7PDG6      | Q7pdg6 clostridium   |
| 70  | 109   | 7.2 | 921  | 2 | Q7PDQ9      | Q7pdq9 fusobacteri   |
| 71  | 109   | 7.2 | 1363 | 2 | Q874Y4      | Q874y4 podospora a   |
| 72  | 109   | 7.2 | 1492 | 2 | Q68724      | Q68724 yersinia pe   |
| 73  | 109   | 7.2 | 1545 | 2 | Q9ZH03      | Q9zh03 yersinia pe   |
| 74  | 109   | 7.2 | 1545 | 2 | Q7ARG4      | Q7arg4 yersinia pe   |
| 75  | 109   | 7.2 | 1545 | 2 | AAS58642    | Aas58642 yersinia    |
| 76  | 108.5 | 7.2 | 852  | 2 | Q8I754      | Q8i754 plasmodium    |
| 77  | 108.5 | 7.2 | 884  | 2 | Q8PW31      | Q8pw31 methanosarc   |
| 78  | 108.5 | 7.2 | 953  | 2 | Q73E75      | Q73e75 bacillus ce   |
| 79  | 108.5 | 7.2 | 953  | 2 | AAS39419    | Aas39419 bacillus    |
| 80  | 108.5 | 7.2 | 1175 | 2 | Q8XNW6      | Q8xnw6 clostridium   |
| 81  | 108.5 | 7.2 | 1273 | 2 | Q9BL02      | Q9bl02 caenorhabdi   |
| 82  | 108.5 | 7.2 | 1959 | 1 | MYH9_CHICK  | P14105 gallus gall   |
| 83  | 108.5 | 7.2 | 1999 | 2 | Q63731      | Q63731 rattus norv   |
| 84  | 108   | 7.2 | 622  | 1 | SR68_CAEEL  | Q20822 caenorhabdi   |
| 85  | 108   | 7.2 | 723  | 2 | Q8F0F9      | Q8f0f9 mesoplasma    |
| 86  | 108   | 7.2 | 792  | 2 | Q7ARQ9      | Q7arq9 oryza sativ   |
| 87  | 108   | 7.2 | 872  | 2 | Q7NFE9      | Q7nfe9 gloeobacter   |
| 88  | 108   | 7.2 | 1877 | 2 | Q6BFD6      | Q6bfd6 paramescium   |
| 89  | 108   | 7.2 | 2097 | 2 | Q7RZF8      | Q7rzf8 neurospora    |
| 90  | 108   | 7.2 | 2115 | 2 | Q6MFH6      | Q6mfh6 neurospora    |
| 91  | 108   | 7.2 | 2115 | 2 | CAE85613    | CAE85613 neurospor   |
| 92  | 107.5 | 7.1 | 534  | 2 | Q6FNN2      | Q6fnn2 homo sapien   |
| 93  | 107.5 | 7.1 | 534  | 2 | AAT01278    | Aat01278 homo sapi   |
| 94  | 107.5 | 7.1 | 720  | 2 | Q9H6Q7      | Q9h6q7 trypanosoma   |
| 95  | 107   | 7.1 | 318  | 2 | Q26749      | Q26749 trypanosoma   |
| 96  | 107   | 7.1 | 1611 | 2 | Q7RD43      | Q7rd43 plasmodium    |
| 97  | 107   | 7.1 | 4368 | 2 | Q61851      | Q61851 caenorhabdi   |
| 98  | 106.5 | 7.1 | 527  | 2 | Q74JAO      | Q74jao lactobacill   |
| 99  | 106.5 | 7.1 | 527  | 2 | AAS09030    | Aas09030 lactobacil  |
| 100 | 106.5 | 7.1 | 1938 | 1 | MYSD_CAEEL  | P02567 caenorhabdi   |

ALIGNMENTS



|  |  |   |
|--|--|---|
| RESULT 1   |  |   |
| HYE SALPA  | STANDARD;  | PRT; 302 AA.  |
| ID   | HYE SALPA  |   |
| AC   | O93R6;   |   |
| DT   | 29-MAR-2004 (Rel. 43, Created)   |   |
| DT   | 29-MAR-2004 (Rel. 43, Last sequence update)  |   |
| DT   | 05-JUL-2004 (Rel. 44, Last annotation update)  |   |
| DE   | Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).  |   |
| GN   | Name=hlyE; Synonyms=clyA, sheA;  |   |
| OS   | Salmonella paratyphi-a   |   |
| OC   | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  |   |
| OC   | Enterobacteriaceae; Salmonella.  |   |
| OX   | NCBI_TaxID=54388;  |   |
| RN   | [1]  |   |
| RP   | SEQUENCE FROM N.A.   |   |
| RC   | STRAIN=SMI S3068/99;   |   |
| RX   | MEDLINE=22215712; PubMed=12228306;   |   |
| RA   | Oscarsson J., Westermarck M., Loefdaahl S., Olsen B., Palmgren H.,   |   |
| RA   | Mizunoe Y., Wai S.N., Uhlin B.E.;  |   |
| RT   | "Characterization of a pore-forming cytotoxin expressed by Salmonella  |   |
| RT   | enterica serovars typhi and paratyphi A.";   |   |
| RL   | Infect. Immun. 70:5759-5769(2002).   |   |
| CC   | -!- FUNCTION: Toxin, which has some hemolytic activity towards   |   |
| CC   | mammalian cells. Acts by forming a pore-like structure upon  |   |
| CC   | contact with mammalian cells (By similarity).  |   |
| CC   | -!- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a  |   |
| CC   | monomer, while in outer membrane vesicles, it oligomerizes to form   |   |
| CC   | a pore structure that is active. Probably forms an octamer (By   |   |
| CC   | similarity).   |   |
| CC   | -!- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer  |   |
| CC   | membrane vesicles. Also found in the periplasmic space (By   |   |
| CC   | similarity).   |   |
| CC   | -!- PTM: In periplasm, it forms a disulfide bond, which prevents the   |   |
| CC   | oligomerization. In outer membrane vesicles, the redox status  |   |
| CC   | prevents formation of the disulfide bond, leading to   |   |
| CC   | oligomerization and pore formation (By similarity).  |   |
| CC   | -!- SIMILARITY: Belongs to the hemolysin E family.   |   |
| CC   | -----  |   |
| CC   | This SWISS-PROT entry is copyright. It is produced through a collaboration   |   |
| CC   | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |   |
| CC   | the European Bioinformatics Institute. There are no restrictions on its  |   |
| CC   | use by non-profit institutions as long as its content is in no way   |   |
| CC   | modified and this statement is not removed. Usage by and for commercial  |   |
| CC   | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |   |
| CC   | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).  |   |
| CC   | -----  |   |
| CC   | EMBL; AJ313033; CAC38362.1; -  |   |
| DR   | HSSP; P77335; 1QOV.  |   |
| DR   | InterPro; IPR010356; HlyE.   |   |
| DR   | Pfam; PF06109; HlyE; 1.  |   |
| KW   | Cytolysis; Hemolysis; Toxin; Transmembrane.  |   |
| FT   | INIT MET 0 By similarity.  |   |
| FT   | TRANSMEM 178 198 Potential.  |   |
| FT   | DISULFID 86 284 In monomeric form (By similarity).   |   |
| SQ   | SEQUENCE 302 AA; 33613 MW; 22DB9112B763A061 CRC64;   |   |
| Query Match  |  |   |
| Best Local Similarity 100.0%; Pred. No. 1.4e-91;             |  |   |
| Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |  |   |
| QY   | 2  | TGIFAQTEVWVKSATETADGALDFYKYLQVLPKTFDETIKELSRFQEQVSEASVL 61    |
| DB   | 1  | TGIFAQTEVWVKSATETADGALDFYKYLQVLPKTFDETIKELSRFQEQVSEASVL 60    |
| QY   | 62   | VGDIKVLMDSDQKYFEATQTVYVWCGVTVQLLSAYILLDFEYNEKKAQKIDILRIID 121 |
| DB   | 61   | VGDIKVLMDSDQKYFEATQTVYVWCGVTVQLLSAYILLDFEYNEKKAQKIDILRIID 120 |
| QY   | 122  | DGVNKLNEAKSLIGSSQFNNSAGKLALDSQITNDFSEKSSYFQSQVDRIKAYAGA 181   |
| DB   | 121  | DGVNKLNEAKSLIGSSQFNNSAGKLALDSQITNDFSEKSSYFQSQVDRIKAYAGA 180   |

|  |  |   |
|--|--|---|
| RESULT 2   |  |   |
| HYE SALTI  | STANDARD;  | PRT; 302 AA.  |
| ID   | HYE SALTI  |   |
| AC   | O8Z727; Q934C4;  |   |
| DT   | 29-MAR-2004 (Rel. 43, Created)   |   |
| DT   | 29-MAR-2004 (Rel. 43, Last sequence update)  |   |
| DT   | 01-OCT-2004 (Rel. 45, Last annotation update)  |   |
| DE   | Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).  |   |
| GN   | Name=hlyE; Synonyms=clyA, sheA; OrderedLocusNames=STV1498, t1477;  |   |
| OS   | Salmonella typhi.  |   |
| OC   | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  |   |
| OC   | Enterobacteriaceae; Salmonella.  |   |
| OX   | NCBI_TaxID=601;  |   |
| RN   | [1]  |   |
| RP   | SEQUENCE FROM N.A.   |   |
| RC   | STRAIN=Ty21a, and SMI S2369/96;  |   |
| RX   | MEDLINE=22215712; PubMed=12228306;   |   |
| RA   | Oscarsson J., Westermarck M., Loefdaahl S., Olsen B., Palmgren H.,   |   |
| RA   | Mizunoe Y., Wai S.N., Uhlin B.E.;  |   |
| RT   | "Characterization of a pore-forming cytotoxin expressed by Salmonella  |   |
| RT   | enterica serovars typhi and paratyphi A.";   |   |
| RL   | Infect. Immun. 70:5759-5769(2002).   |   |
| CC   | -!- FUNCTION: Toxin, which has some hemolytic activity towards   |   |
| CC   | mammalian cells. Acts by forming a pore-like structure upon  |   |
| CC   | contact with mammalian cells (By similarity).  |   |
| CC   | -!- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a  |   |
| CC   | monomer, while in outer membrane vesicles, it oligomerizes to form   |   |
| CC   | a pore structure that is active. Probably forms an octamer (By   |   |
| CC   | similarity).   |   |
| CC   | -!- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer  |   |
| CC   | membrane vesicles. Also found in the periplasmic space (By   |   |
| CC   | similarity).   |   |
| CC   | -!- PTM: In periplasm, it forms a disulfide bond, which prevents the   |   |
| CC   | oligomerization. In outer membrane vesicles, the redox status  |   |
| CC   | prevents formation of the disulfide bond, leading to   |   |
| CC   | oligomerization and pore formation (By similarity).  |   |
| CC   | -!- SIMILARITY: Belongs to the hemolysin E family.   |   |
| CC   | -----  |   |
| CC   | This SWISS-PROT entry is copyright. It is produced through a collaboration   |   |
| CC   | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |   |
| CC   | the European Bioinformatics Institute. There are no restrictions on its  |   |
| CC   | use by non-profit institutions as long as its content is in no way   |   |
| CC   | modified and this statement is not removed. Usage by and for commercial  |   |
| CC   | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |   |
| CC   | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).  |   |
| CC   | -----  |   |
| CC   | EMBL; AJ313033; CAC38362.1; -  |   |
| DR   | HSSP; P77335; 1QOV.  |   |
| DR   | InterPro; IPR010356; HlyE.   |   |
| DR   | Pfam; PF06109; HlyE; 1.  |   |
| KW   | Cytolysis; Hemolysis; Toxin; Transmembrane.  |   |
| FT   | INIT MET 0 By similarity.  |   |
| FT   | TRANSMEM 178 198 Potential.  |   |
| FT   | DISULFID 86 284 In monomeric form (By similarity).   |   |
| SQ   | SEQUENCE 302 AA; 33613 MW; 22DB9112B763A061 CRC64;   |   |
| Query Match  |  |   |
| Best Local Similarity 100.0%; Pred. No. 1.4e-91;             |  |   |
| Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |  |   |
| QY   | 2  | TGIFAQTEVWVKSATETADGALDFYKYLQVLPKTFDETIKELSRFQEQVSEASVL 61    |
| DB   | 1  | TGIFAQTEVWVKSATETADGALDFYKYLQVLPKTFDETIKELSRFQEQVSEASVL 60    |
| QY   | 62   | VGDIKVLMDSDQKYFEATQTVYVWCGVTVQLLSAYILLDFEYNEKKAQKIDILRIID 121 |
| DB   | 61   | VGDIKVLMDSDQKYFEATQTVYVWCGVTVQLLSAYILLDFEYNEKKAQKIDILRIID 120 |
| QY   | 122  | DGVNKLNEAKSLIGSSQFNNSAGKLALDSQITNDFSEKSSYFQSQVDRIKAYAGA 181   |
| DB   | 121  | DGVNKLNEAKSLIGSSQFNNSAGKLALDSQITNDFSEKSSYFQSQVDRIKAYAGA 180   |



```

CC  -!- SIMILARITY: Belongs to the hemolysin E family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AJ313032; CAC38360.1; -
DR  EMBL; AJ313034; CAC38363.1; -
DR  EMBL; AE627270; CAD01758.1; ALT_INIT.
DR  EMBL; AE016839; AAO69115.1; ALT_INIT.
DR  HSP; P77335; 100Y.
DR  InterPro; IPR010356; HlyE.
DR  Pfam; PF06109; HlyE; 1
KW  Complete proteome; Cytolysis; Hemolysis; Toxin; Transmembrane.
FT  INIT_MET 0 By similarity.
FT  TRANSMEM 178 198 Potential.
FT  DISULFID 86 284 In monomeric form (By similarity).
SQ  SEQUENCE 302 AA; 33658 MW; 2BDFD835D044FDAE CRC64;

Query Match 97.2%; Score 1467; DB 1; Length 302;
Best Local Similarity 97.4%; Pred. No. 4.6e-89;
Matches 294; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 2 TGIFAEQVEVVKSAIETADGALDPYKYLQVLPWKTPDETKELSRFKQYSEASVL 61
Db 1 TGIFAEQVEVVKSAIETADGALDPYKYLQVLPWKTPDETKELSRFKQYSEASVL 60

Qy 62 VGDIKVLLMDSQDKYFEATQTVYECGVVTVQLLSAYILLFDEYNKKASAKDILIRILD 121
Db 61 VGDIKVLLMDSQDKYFEATQTVYECGVVTVQLLSAYILLFDEYNKKASAKDILIRILD 120

Qy 122 DVNKLNAQKSLIGSSQSFNNASGKLLALDSQLTNDSEKSSYPQSDVRIKRAYAGA 181
Db 121 DVNKLNAQKSLITSSQSFNNASGKLLALDSQLTNDSEKSSYPQSDVRIKRAYAGA 180

Qy 182 AAGIVAGPFLGIISYIAAGVIEGKLIPELNLKAVONFTSLSVTVKQAKDIDAAL 241
Db 181 AAGIVAGPFLGIISYIAAGVIEGKLIPELNLKATVQNFSTLSATVQAKDIDAAL 240

Qy 242 KLATEIAAIGIKETETETTRFYVDYDMLSLKGAKKMINTCNEYQQRHGKTLLEVP 301
Db 241 KLATEIAAIGIKETETETTRFYVDYDMLSLKGAKKMINTCNEYQQRHGKTLLEVP 300

Qy 302 DI 303
Db 301 DV 302

RESULT 3
HLYE_ECOLI STANDARD; PRT; 302 AA.
AC P77335; Q47276; Q8VU70; Q9R3G4;
DT 01-NOV-1997 (Rel. 35, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Hemolysin E, chromosomal (hemolysis-inducing protein) (Silent
DE hemolysin sheA) (Cytotoxin clyA) (latent pore-forming 34 kDa
DE haemolysin).
DN Name=hlyE; Synonyms=clyA, sheA, hpr; OrderedLocusNames=bl182;
OS Escherichia coli
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / XLI-BLUE;
RA McNamara P.J., Iandolo J.J., Uhlich G.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA del Castillo F.J., Leal S.C., Moreno F., del Castillo I.;
RT "The Escherichia coli K-12 sheA gene encodes a 34-kDa secreted
RT haemolysin."
RL Mol. Microbiol. 25:107-115(1997).
[3]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-12, SUBCELLULAR LOCATION, AND
RP INDUCTION.
RC STRAIN=K12;
EX MEDLINE=99157562; PubMed=10027972;
RA Ludwig A., Bauer S., Benz R., Bergmann B., Goebel W.;
RT "Analysis of the SlyA-controlled expression, subcellular localization
RT and pore-forming activity of a 34 kDa haemolysin (ClyA) from
RT Escherichia coli K-12."
RL Mol. Microbiol. 31:557-567(1999).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
EX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
EX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
[7]
RP SEQUENCE FROM N.A.
RC STRAIN=CH9802;
RA Chang G.-N., Ho K.-C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[8]
RP SEQUENCE OF 1-295 FROM N.A.
RC STRAIN=K12 / XLI-BLUE;
RA King C.H., Shinnick T.M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
[9]
RP SEQUENCE OF 1-155 FROM N.A.
RC STRAIN=K12 / AB1157;
RA Woodgate R.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
[10]
RP SEQUENCE OF 1-17, AND MUTAGENESIS OF 87-GLY--VAL-89; 142-ASN-ALA-143;
RP 182-ALA-GLY-183; 186-ALA-GLY-187; ASP-267 AND 292-GLY-LYS-293.
RX MEDLINE=99316011; PubMed=10383763;
RA Oscarson J., Mizunoe Y., Li L., Lai X.-H., Wieslander A., Uhlin B.E.;
RT "Molecular analysis of the cytolytic protein ClyA (SheA) from
RT Escherichia coli."
RL Mol. Microbiol. 32:1226-1238(1999).
[11]
RP PARTIAL SEQUENCE, FUNCTION, SUBCELLULAR LOCATION, OLIGOMERIZATION, AND
RP DISULFIDE BOND FORMATION.
EX MEDLINE=22894281; PubMed=14532000;
RA Wal S.N., Lindmark B., Soederblom T., Takade A., Westermark M.,
RA Oscarson J., Jass J., Richter-Dahlfors A., Mizunoe Y., Uhlin B.E.;

```



GN Name=hlyE; Synonyms=clyA, sheA; OrderedLocusNames=1944, ECS1677;  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EHEC;  
RX MEDLINE=20322319; PubMed=10865950; DOI=10.1016/S0923-2508(00)00143-1;  
RA del Castillo F.J., Moreno F., del Castillo I.;  
RT "Characterization of the genes encoding the SheA haemolysin in  
Escherichia coli O157:H7 and Shigella flexneri 2a";  
RL Res. Microbiol. 151:229-230(2000).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;  
RX MEDLINE=21074915; PubMed=11206551; DOI=10.1038/35054089;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533(2001).  
RN (3)  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:11-22(2001).  
CC -!- FUNCTION: Toxin, which has some hemolytic activity towards  
mammalian cells. Acts by forming a pore-like structure upon  
contact with mammalian cells (By similarity).  
CC -!- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a  
monomer, while in outer membrane vesicles, it oligomerizes to form  
a pore structure that is active. Probably forms an octamer (By  
similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer  
membrane vesicles. Also found in the periplasmic space (By  
similarity).  
CC -!- PTM: In periplasm, it forms a disulfide bond, which prevents the  
oligomerization. In outer membrane vesicles, the redox status  
prevents formation of the disulfide bond, leading to  
oligomerization and pore formation (By similarity).  
CC -!- SIMILARITY: Belongs to the hemolysin E family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; AJ238954; CAB64962.1; ALT\_INIT.  
CC EMBL; AF005335; AAG56033.1; ALT\_INIT.  
CC EMBL; AF002555; BAB35100.1; ALT\_INIT.  
CC PIR; E85696; E85696.  
CC PIR; E90838; E90838.  
CC HSP; P77335; 1Q0Y.  
CC InterPro; IPR010356; HlyE.  
CC Pfam; PF06109; HlyE; 1.  
CC Complete proteome; Cytolysis; Hemolysis; Toxin; Transmembrane.  
FT INIT MET 0 0 By similarity.  
FT TRANSMEM 182 202 Potential.  
FT DISULFID 86 284 In monomeric form (By similarity).  
SQ SEQUENCE 302 AA; 33585 MW; F261E29E1DE5FC87 CRC64;

Query Match 90.1%; Score 1361; DB 1; Length 302;  
Best Local Similarity 88.7%; Pred. No. 4.6e-82;  
Matches 268; Conservative 20; Mismatches 14; Indels 0; Gaps 0;  
QY 2 TGIFAEQTVVVKSAIETADGALDFYNYKLDQVLPWKTFDETIKLSRPFQKQYSQASV 61  
DB 1 TEIVADKTVVVKNAIETADGALDLYNKYLDQVLPWKTFDETIKLSRPFQKQYSQA 60  
QY 62 VGDIKVLMDSDQKYFEATQTVYEWCVVTVLLSAYILLFDYNEKKASAKDILIRLD 121  
DB 61 VGNIKVLMDSDQKYFEATQTVYEWCVVTVLLSAYILLFDYNEKKASAKDILIRLD 120  
QY 122 DGVNKLNEAQSLLGSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKREYAGA 181  
DB 121 DGITKLENAQSKLLVSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKREYAGA 180  
QY 182 AAGIVAGPFLIISYIAAGVIEGKLIPELANDRLKAVQNFETSLVTVKQAKNDIDAKL 241  
DB 181 AAGVAGPFLIISYIAAGVIEGKLIPELNKLSQVQSFETSLVTVKQAKNDIDAKL 240  
QY 242 KLATEIAAIGIKETETETTRFVYDDYDMLSLKGAKKMINTCNEYQORHGKKTLLFVP 301  
DB 241 KLTETIAAIGIKETETETTRFVYDDYDMLSLKGAKKMINTCNEYQORHGKKTLLFVP 300  
QY 302 DI 303  
DB 301 EV 302  
RESULT 5  
QX2S8 PRELIMINARY; PRT; 300 AA.  
ID Q9X2S8  
AC Q9X2S8;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Hemolysin.  
GN Name=hlyE;  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=99242013; PubMed=10227474;  
RX MEDLINE=99242013; PubMed=10227474;  
RA Reingold J., Starr N., Maurer J., Lee M.D.;  
RT "Identification of a new Escherichia coli She haemolysin homolog in  
avian B. coli";  
RL Vet. Microbiol. 66:125-134(1999).  
DR EMBL; AF052225; AAD28079.1; -.  
DR HSP; P77335; 1Q0Y.  
DR InterPro; IPR010356; HlyE.  
DR Pfam; PF06109; HlyE; 1.  
SQ SEQUENCE 300 AA; 33555 MW; 6580B66C44A7B4BC CRC64;  
Query Match 73.1%; Score 1104; DB 2; Length 300;  
Best Local Similarity 73.4%; Pred. No. 4.5e-65;  
Matches 213; Conservative 39; Mismatches 38; Indels 0; Gaps 0;  
QY 6 AEQTVVVKSAIETADGALDFYNYKLDQVLPWKTFDETIKLSRPFQKQYSQASV 65  
DB 4 ADQTVETVKTATDADKALDLYNKYLDQVLPWKTFDETIKLSRPFQKQYSQASV 63  
QY 66 KVLMDSDQKYFEATQTVYEWCVVTVLLSAYILLFDYNEKKASAKDILIRLDGWN 125  
DB 64 KSLMNSQDRYFEATQTVYEWCVVTVLLSAYILLFDYNEKKASAKDILIRLDGWN 123  
QY 126 KLENAQSKLLGSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKREYAGA 185  
DB 124 KLENAQSKLLGSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKREYAGA 183



```

QY 211 LNDRLKAVQNTFTSLVTVKQAKDIDAAKLKLAIEAIGIKTETETTRFYVDYDDLM 270
DB 1 MNKLKLSALSPFTLSTVTVKQAKDIDAAKLKLAIEAIGIKTETETTRFYVDYDDLM 60

QY 271 LSLKGAAGKMTNCNEVQORHGKTKLLRPDI 303
DB 61 LSLKGAAGKMTNCNEVQORHGKTKLLRPDI 93

RESULT 8
Q9LIW7 PRELIMINARY; PRT; 895 AA.
AC Q9LIW7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Similar to an Arabidopsis thaliana chromosome BAC genomic
DE sequence.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RL Hsing Y.C., Chow T., Chen C., Wu H., Chu M., Chao Y., Liu S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001111; BAA90502.1; -.
DR Gramene; Q9LIW7; -.
SQ SEQUENCE 895 AA; 100520 MW; BA4F7EFD70FFA3 CRC64;

Query Match 9.0%; Score 136.5; DB 2; Length 895;
Best Local Similarity 22.4%; Pred. No. 1.6;
Matches 77; Conservative 67; Mismatches 121; Indels 79; Gaps 17;

QY 8 QTVVVKGAIEATDAGD-----FYNKYLQDVIPWTKTDFDTIK-----ELSRFKQY 54
DB 274 ERVEITLSSEVRLKGLDSTAESSESKNRETELV--KNLESEVSLVKGLFEARIIER 331

QY 55 SOEASVLGCDIKVLLMDSQDKYFEATQTVYEW---CGVVQTLLSAVILLFQY-NEKAS 110
DB 332 LAETEKLEELKSEVADAKAEARQIFEEWKHAGLLEMELEA-VTLDKFKGSELAS 390

QY 111 AOKDILIRILDGVNKLNEAOKSLGSSOSFNASGKLIALLD---SQTNDPFSKSYF- 166
DB 391 TTEE-----LGKIQSALQDRSEIEVLKGTITALEIEVARLLADVNESNEQFD 438

QY 167 -----QSQVDRIRKEAYAGAAIVAGPFGGLIISYSIAAGV-IEG-----KLPIE 210
DB 439 ASQOEVFGLQTTIDVLRNKLNEAEEAASEA-----LNNEKAANYKIEGLTEENVKLISE 492

QY 211 LND-----RLKAVQNTFTSLVTVKQAKDIDAAKLKLAIEAIGIKTET 257
DB 493 LNETDRBEKEKRAVEDLTAALUS-----ESSDKAKEAHERVLSKEDDHEHALAIGOLKMAL 549

QY 258 ETTR----FYVDYDDLMLSLLKGAAGKMTNCNEVQOR-HGKKT 296
DB 550 KSTKSEYEWLDEANYDITCLRNKVDKLEAEVKNKYRESCEKSET 593

RESULT 9
P71497 PRELIMINARY; PRT; 495 AA.
AC P71497;
DT 01-FEB-1997 (TRENBLrel. 02, Created)
DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE ClpB (fragment).
GN Name=clpB;
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2095;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25416;
RX MEDLINE=971148974; PubMed=8995799;
RA Falah M., Gupta R.S.;
RT "Phylogenetic analysis of mycoplasmas based on Hsp70 sequences:
RT cloning of the dnaK (hsp70) gene region of Mycoplasma capricolum.";
RL Int. J. Syst. Bacteriol. 47:38-45(1997).
DR EMBL; U51235; AAB09427.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001270; Chaperin_c1pA/B.
DR PRINTS; PRO0300; CLPPTROTEASEA.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00871; CLPAB_2; 1.
KW ATP-binding.
FT NON_TER 1
SQ SEQUENCE 495 AA; 56358 MW; A194DD51FACDF8D0 CRC64;

Query Match 8.7%; Score 132; DB 2; Length 495;
Best Local Similarity 24.3%; Pred. No. 1.5;
Matches 79; Conservative 47; Mismatches 115; Indels 84; Gaps 16;

QY 2 TGIFAEQTVVVKGAIEATDAGDGFYNKYLQDVIP-----WKTDFETIKELSR 49
DB 178 TGIIVDKLISSEKERLLNLEDLLKKYVKGQDAIKAVTSAIMRSRGIKNPKPIGSEFLF 237

QY 50 F-----KQEYSQE-ASVLVGDIKVLLMDSQDKYFEATQTV-----YEWCGVVT 91
DB 238 FGPTGVGKTEVARSLADILFNSPKMIRLDMSEYMEKHSVAKLIGAPPGYGVYEGGRLLT 297

QY 92 QLL--SAY-ILLFQYNEKKASQAKDILIRILDGVNKLNEAOKSLGSSOSFNASGKL 148
DB 298 EAVRNPYSIILFDEI-EKAHSDVFNILLOILDDG--RLTDS-----LGKTIIDFKNT---I 347

QY 149 LALDSQLTNDPFSKSYF-----QSQVDRIRKEAYAGAAAGIVAGPFGGLIISYSIAA 200
DB 348 IVMTSNIAQYLLTSDDDFVQDDQKIQAEIQLNLTQTFRPEFLNRIDNIVY--FNALSQVTI-- 403

QY 201 GVIEGKLIPENDRLKAVQNTFTSLVTVKQAKDIDAAKLKLAIEAIGIKTETET 260
DB 404 GEIVDKLDELITRLQDQEQNYFINPS-----EARNKIINE----- 439

QY 261 RFYVDYDDLM--LSLLKGAAGKMTNC 284
DB 440 ---CYDFLFGARLLKDLIKLIET 460

RESULT 10
Q7ROB6 PRELIMINARY; PRT; 2723 AA.
AC Q7ROB6;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE 235 kDa rhostry protein.
GN Name=py01185;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteza M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,

```

RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J., Carucci D.J.;  
RA "Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii";  
RT Nature 419:512-519(2002). shown here is derived from an  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.  
CC EMBL: AABL01000312; EAA20475.1; -;  
DR InterPro: IPR011561; Pox 11 rel.  
DR InterPro: IPR006499; ReticulocyteP.  
DR ProDom: PD014111; Pox 11 rel. 1.  
DR TIGRFAM: TIGR01612; 235kDa-fam; 1.  
SQ SEQUENCE 2723 AA; 320451 MW; 08CD9CE7672ED2 CRC64;

Query Match 8.6%; Score 130; DB 2; Length 2723;  
Best Local Similarity 19.7%; Pred. No. 16;  
Matches 62; Conservative 54; Mismatches 96; Indels 102; Gaps 10;

QY 11 EVVKSATETADGALDFYKYLDOVLPWKTFDTIKELSRFKQE----- 53  
Db 1464 EELKGYDKSKGYKDEVNKNT-----KTIENKKELEKYKNEAINLLNKFSEFMLKNKI 1517

QY 54 --YSQASVLDGDIKVLMDSDQKYFEATQTVYVWCVVTVQLLSAYILLFDEYNEKKASA 111  
Db 1518 IOTKKDELIINEIKWKKES-----ILQADKSEQKINT 1551

QY 112 QKDILIRLDDGVNKLNEAQSLLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVD 171  
Db 1552 IKNEQIQKDDTTSN-DKSNKAIIGIQASLDKFKFLKI-----NDIRTKSNFLKETE 1605

QY 172 RIRKEAYAGAAGIVAGPGLIISYSIAGVIEGKLIPELNDRLKAVQNFSTLSVTVKQ 231  
Db 1606 NIEKQ-----ISNLSINSQEIK-----LKNEDILNTLQKFESL-----KD 1642

QY 232 ANKQIDAACKLKLAT-----ETAAIGEIKETETETTFYVYDYLMLLS 272  
Db 1643 QKQVEDQKTELDFDSEIENIESDVSNQKNVEMGIEIKENADTNKQIESTK---E 1699

QY 273 LKGAAGKMNNTON 286  
Db 1700 LIRPTIENLISFN 1713

RESULT 11  
Q6MT03 PRELIMINARY; PRT; 713 AA.  
AC Q6MT03;  
DT 05-JUL-2004 (T-EMBLrel. 27, Created)  
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
DE ATP dependant protease ClpB.  
GN Name=clpB; OrderedLocusNames=MSC\_0613;  
OS Mycoplasma mycoides (subsp. mycoides SC).  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=44101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PG1;  
RX PubMed=14762060;  
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundeberg J., Johansson K.-E., Pettersson B., Uhlen M.;  
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP).";  
RT Genome Res. 14:221-227(2004).  
RL EMBL: BX842644; CAE77235.1; -;  
DR EMBL; BX842644; CAE77235.1; -;  
GO: GO:0008233; F:peptidase activity; IEA.  
DR InterPro: IPR003593; AAA ATPase.  
DR InterPro: IPR003959; AAA ATPase centr.  
DR InterPro: IPR001270; Chaprinin\_c1pA/B.  
DR Pfam: PF00004; AAA; 2  
DR PRINTS: PR00300; CLPPTASEA.

DR SMART: SM00382; AAA; 2.  
DR PROSITE: PS00871; CLPAB\_2; 1.  
KW ATP-binding; Complete proteome; Protease.  
SQ SEQUENCE 713 AA; 80782 MW; 1B5D204A9E29BE50 CRC64;

Query Match 8.5%; Score 128.5; DB 2; Length 713;  
Best Local Similarity 24.5%; Pred. No. 4;  
Matches 81; Conservative 48; Mismatches 124; Indels 77; Gaps 17;

QY 2 TGIFAEQTVVVKSAIETADGALDFYKYLDOVLPWKTFDTI-----KELSRF 50  
Db 396 TGIIVDRLLISSEKEKLNLEDLLKKYVKGQDQAI--KAVTSALMRSGIKNPKDPGIFGSF 453

QY 51 -----KQYISOE-ASVLVGDIKVLMDSDQKYFEATQTV-----YEWCGV 89  
Db 454 LFLGPTGVGKTEVARSLADILFNSPKMIRLDMSEYMEKHSVAKLIGAPPVGVYEGGR 513

QY 90 VTQLL--SAY-ILLFDEYNEKKASAQKDIILIRLDDGVNKLNEAQSLLGSSQSFNN--- 143  
Db 514 LTAVERNRPYSIVLFDEI-EKAHTDVFNILLQILDG--RLTDS-----LGKTTDFKNTII 566

QY 144 -----ASGKLLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAGAAAGIVAGPGLIISYS 197  
Db 567 VMTSNIASQVLLTSDELQVD-----DQIQEELNKVFRPEFLNRIDNIVY--FNALSQVT 620

QY 198 IAAGVIEGKLIPELNDRLKAVQNFSTLSVTVKQANKDIDAACKL--ATEIAAIGEIKT 255  
Db 621 I--GEIVDKVLEELSTRLQDEQNVFINFSEAR--NKIINEGYDRLFGARPIKRYIEKNI 676

QY 256 ET-----ETTRFYVDYDMLLSL 273  
Db 677 ETLIAHYIISGEVVVENTRYLIDVKNNOFTL 706

RESULT 12  
CAE77235 PRELIMINARY; PRT; 713 AA.  
AC CAE77235;  
DT 02-MAR-2004 (T-EMBLrel. 27, Created)  
DT 02-MAR-2004 (T-EMBLrel. 27, Last sequence update)  
DT 13-APR-2004 (T-EMBLrel. 27, Last annotation update)  
DE ATP dependant protease ClpB.  
GN CLPB OR MSC\_0613.  
OS Mycoplasma mycoides (subsp. mycoides SC).  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=44101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PG1;  
RX PubMed=14762060;  
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundeberg J., Johansson K.-E., Pettersson B., Uhlen M.;  
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP).";  
RT Genome Res. 14:221-227(2004).  
RL EMBL: BX842644; CAE77235.1; -;  
DR EMBL; BX842644; CAE77235.1; -;  
GO: GO:0008233; F:peptidase activity; IEA.  
DR InterPro: IPR003593; AAA ATPase.  
DR InterPro: IPR003959; AAA ATPase centr.  
DR InterPro: IPR001270; Chaprinin\_c1pA/B.  
DR Pfam: PF00004; AAA; 2  
DR PRINTS: PR00300; CLPPTASEA.

```

Db 514 LTEAVRNYSIVLPEDEI-EKAHTDVFNILLQLDDG--RLTDS-----LGKTIIDFKNTII 566
Qy 144 -----ASGKLALDSQLTNDSEKSSVFQSDVRIRKEAYAGAAAGIVAGPFGILIIISVS 197
Db 567 VMTSNIASQYLTSBELVQVD-----DQKIQEELNKVFRPEFLNRIDNIVY--FNALSQVT 620
Qy 198 IAAGVIEGKLIPELNDRKAVQNFTSISVTVKQANKOIDAAKJL--ATEIAAIGEIKT 255
Db 621 I--GEIVDKVLELSTRLQDEQNYFINFSEEAR--NKIINEGYDRLFGARPIKRYIEKNI 676
Qy 256 ET-----ETTFYVDYDDLMLSL 273
Db 677 ETLIAHYIISGEVVENTRYLIDVKNQNFYL 706

RESULT 13
Q7N5U5 PRELIMINARY; PRT; 577 AA.
AC Q7N5U5;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Similar to putative phage protein.
GN OrderedLocNames=plu1837;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314;
RA Duchaud E., Rusnok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-P.,
RA Dassa E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,
RA Medigue C., Lanolis A., Powell K., Sigulier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens."
RL Nat. Biotechnol. 21:1307-1313 (2003).
DR EMBL; BX571865; CAE14130.1; -.
DR PhotocList; plu1837; -.
KW Complete proteome.
SQ SEQUENCE 577 AA; 65287 MW; B7F86BE4B34BEDE1 CRC64;

Query Match 8.4%; Score 126.5; DB 2; Length 577;
Best Local Similarity 19.0%; Pred. No. 4.2; Mismatches 155; Gaps 14;
Matches 80; Conservative 55; Indels 132; Indels 155; Gaps 14;

Qy 1 MTGIPAEQTVGVK-----SAIETADGALDFYKYLQV-----IPWKTF----- 40
Db 103 ITGDSKGFSDLVNFEMKLPNREGDNLTPPPAYVFLPFIQIKSWSPWNSFENLQY 162
Qy 41 -----DETIKELSRFKQYSQEASVLVDIKVL----- 68
Db 163 SNWKGPIKYFTGLKPEHFOEENIYEYSEIKESAHKIEKFSQAVEIVDNTVDSPT 222
Qy 69 -----LMSQDKYPEATQTVYEMGVVTVLLSAYILLFDEYNEKKA 109
Db 223 LONNDFIKIONEIKNELYDLIDYQRLYDAQTI--TSNIYDLESQVALAITSANELEA 279
Qy 110 -----SAQKDL-----IRILDDGVNKLNEAQKSLGSSQSPNNASG 146
Db 280 DYKFAVESIPTDYLECPLCGTLHDNSLPNRLALLSEKSLINEA-----NSIAS 328
Qy 147 KLLALDSQLTNDFSEKSSVFQSDVRIRKEAYAGAAAGIVAGPFGILIIISYIAAGVIEGK 206
Db 329 KIVELKSSL-NSLNEDAQFITNEIERINKK-----YITDDAVKKG 368
Qy 207 LIPELNDRL-----KAVQNFTSISVTVKQANKDI-----DAAKL----- 241
Db 369 LIAQVIDTLSTENVSKNTQIKIDNEDLNIISKANSISKELKKDQKLLSTYKEELNSSF 428
```

```

Qy 242 -KLATEIAAIGEIKTETETTRFYVDYDDLMLSLLLKGAARKMI-NTCNEYQQRHGKKTLL 299
Db 429 SKLLANIEALSGTGINLSKVSPDQYKQLLGGAAGAAAGLLAYQLSVLQIHGAKTCVV 488
Qy 300 VP 301
Db 489 PP 490

RESULT 14
QSG015 PRELIMINARY; PRT; 1521 AA.
AC QSG015;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocNames=BQ05330;
OS Bartonella quintana (Rochalimaea quintana).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=803;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Toulouse;
RX PubMed=15210978;
RA Almark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardell D.H.,
RA Canbaeck B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huvet M.,
RA La Scola B., Holmberg M., Andersson S.G.E.;
RT "The louse-borne human pathogen Bartonella quintana is a genomic
RT derivative of the zoonotic agent Bartonella henselae."
RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721 (2004).
DR EMBL; BX897700; CAF26028.1; -.
DR InterPro; IPR009074; Apolipo_A_E_C3.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR010989; t-snare.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1521 AA; 169742 MW; 12CC93BCB2C97920 CRC64;

Query Match 8.3%; Score 125; DB 2; Length 1521;
Best Local Similarity 18.4%; Pred. No. 17; Mismatches 147; Indels 90; Gaps 8;
Matches 69; Conservative 69; Indels 147; Indels 90; Gaps 8;

Qy 4 IFAEQTVGVKSAIETADGALDFYK-----YLDQVIPWK-----TFDET 43
Db 928 VLSEQTACTVESFTASHNAQTLNETHTSATAIEEVNERNVNLHSMQNLKNLGVQ 987
Qy 44 IKELSRFKQYSQEASVLVDIKVLLMDSQDKYPEATQTVYEMGVVTVLLSAYILLFDE 103
Db 988 LSDVSNHLEEKQTATQISGHVEKLTETLVNLAQAQNTTESISHLTQHIHQISEQLSLSTQ 1047
Qy 104 YNEKASAKQDILIRIL-----DDGVNKLNEAQKSLGSSQSPNNASGKL 148
Db 1048 DAQRIVYAQNESLNSLTQTNTSETLQVTAMKEDLVNNISSILKQLNQSISYFHSNML 1107
Qy 149 LALDSQLTNDSEKSSVFQSDVRIRKEAYAGAAAGIVAGPFGILIIISYIAAGVIE----- 204
Db 1108 LSTVQNDIQSEFANNFRFTNQ-----AAHLASNAQNLNNVTVLQGLSQNLFE 1159
Qy 205 -----GKLIPELNDRLKAVQNFTSISVTVKQANKOIDAAKJLKLATEIAAIGEI-- 253
Db 1160 KIGHTSTFGSHAKTLSETIHILEKSENSLTTLLEKHKTLKSALLVSALVSKSNEINKLIE 1219
Qy 254 -----KTETET-----TFYVDYDDLMLSL--LKGAAKMM 281
Db 1220 YYENVLSLAFERTTNTNRNTHSLQQSLNQLINEASTRFGAAEDIRKSADEIRSELSKI 1279
Qy 282 INTCNEYQQRHGKKT 296
Db 1280 NNDINESVQNLPEKT 1294
```



|  |   |               |  |
|--|---|---------------|--|
| RESULT 15  |   | PRT; 1825 AA. |  |
| Q7RIT2   | PRELIMINARY;  |               |  |
| AC   | Q7RIT2;   |               |  |
| DT   | 01-MAR-2004 (Tremblrel. 26, Created)                                    |               |  |
| DT   | 01-MAR-2004 (Tremblrel. 26, Last sequence update)                       |               |  |
| DE   | Rhoptry protein.  |               |  |
| GN   | Name=PY03534;   |               |  |
| OS   | Plasmodium yoelii yoelii.   |               |  |
| OC   | Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.           |               |  |
| OX   | NCBI_TaxID=73239;   |               |  |
| RN   | [1]   |               |  |
| RP   | SEQUENCE FROM N.A.  |               |  |
| RC   | STRAIN=17XNL;   |               |  |
| RX   | PubMed=12368865;  |               |  |
| RA   | Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteu M.,           |               |  |
| RA   | Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,        |               |  |
| RA   | Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,         |               |  |
| RA   | Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,           |               |  |
| RA   | Cho J.K., Quackenbush J., Sedegah M., Shoabli A., Cummings L.M.,        |               |  |
| RA   | Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,       |               |  |
| RA   | Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,               |               |  |
| RA   | van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,          |               |  |
| RA   | Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,    |               |  |
| RA   | Carucci D.J.;   |               |  |
| RT   | "Genome sequence and comparative analysis of the model rodent malaria   |               |  |
| RT   | parasite Plasmodium yoelii yoelii."                                     |               |  |
| RL   | Nature 419:512-519(2002).   |               |  |
| CC   | -!- CAUTION: The sequence shown here is derived from an                 |               |  |
| CC   | EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is              |               |  |
| CC   | preliminary data.   |               |  |
| DR   | EMBL; AABL01001024; EAA15256.1; -                                       |               |  |
| DR   | InterPro; IPR006499; ReticulocyteBP.                                    |               |  |
| DR   | TIGRFAMs; TIGR01612; 235kDa-fam; 1.                                     |               |  |
| SQ   | SEQUENCE 1825 AA; 213817 MW; 2956D81596385940 CRC64;                    |               |  |
| Query Match 8.2%; Score 124.5; DB 2; Length 1825;                |   |               |  |
| Best Local Similarity 21.1%; Pred. No. 23;                       |   |               |  |
| Matches 61; Conservative 56; Mismatches 105; Indels 67; Gaps 11; |   |               |  |
| QY   | 22 GALDFYKYLQVDPWKTDETIKELSRFKQYISQ---EASVLVGDIKVLLMDSQDKYPE 78         |               |  |
| DB   | 570 GYIDSKSEYKDEV-----DKNTKTIENKKEFEQYKDEAINLLKFSMLKN-----N 618         |               |  |
| QY   | 79 ATQTVVCGVTVQLLSAY--ILLPEYNEKKAQKADILIRLDGVNKLNEAKSLIG 136            |               |  |
| DB   | 619 IIQTKTDLRIIINEIKMKHEKIMQAKSEQKINTIKNEQIQIKDDTTNN--DKSNKALIG 677     |               |  |
| QY   | 137 SSQSFNNASKLALDSQLTNDPSEKSSYFQSQVDRIKEAYAGAAGIVAGPFGLIISY 196        |               |  |
| DB   | 678 IQASLDKPEFKFKI-----NDIRKSNFLKETEINIKO-----ISNL 716                  |               |  |
| QY   | 197 SIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVKQAKDIDAAKLKLTAT----- 245           |               |  |
| DB   | 717 SINS---QGILKDNEDILNTLQTFLESL-----KDQKNIYEQKTELDNFDSEIENIENN 769     |               |  |
| QY   | 246 -----ETAAIGEIKETETETFRFVYDDLMLSLLKGAAKMINTCN 286                    |               |  |
| DB   | 770 INHNKYNENGIEIKENADTNKQIESTK---ELIKPTIENIIISFN 815                   |               |  |
| RESULT 16  |   |               |  |
| Q7SHZ4   | PRELIMINARY;  | PRT; 4007 AA. |  |
| AC   | Q7SHZ4;   |               |  |
| DT   | 01-MAR-2004 (Tremblrel. 26, Created)                                    |               |  |
| DT   | 01-MAR-2004 (Tremblrel. 26, Last sequence update)                       |               |  |
| DE   | Hypothetical protein.   |               |  |
| GN   | Name=NCU00658.1;  |               |  |
| OS   | Neurospora crassa.  |               |  |
| OC   | Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;          |               |  |
| Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.        |   |               |  |
| OX   | NCBI_TaxID=5141;  |               |  |
| RN   | [1]   |               |  |
| RP   | SEQUENCE FROM N.A.  |               |  |
| RC   | STRAIN=OR74A;   |               |  |
| RA   | Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,       |               |  |
| RA   | Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,     |               |  |
| RA   | Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,    |               |  |
| RA   | Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,              |               |  |
| RA   | Selitermnikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,    |               |  |
| RA   | Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,    |               |  |
| RA   | Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,         |               |  |
| RA   | Kamal M., Kamvysselis M., Mauceli E., Bielek C., Rudd S., Frisman D.,   |               |  |
| RA   | Krystofova S., Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S., |               |  |
| RA   | Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,     |               |  |
| RA   | DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,           |               |  |
| RA   | Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,    |               |  |
| RA   | Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,          |               |  |
| RA   | Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,            |               |  |
| RT   | "The Genome Sequence of the Filamentous Fungus Neurospora crassa."      |               |  |
| RL   | Nature 0:0-0(2003).   |               |  |
| CC   | -!- CAUTION: The sequence shown here is derived from an                 |               |  |
| CC   | EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is              |               |  |
| CC   | preliminary data.   |               |  |
| DR   | EMBL; AABX01000001; EAA36562.1; -                                       |               |  |
| DR   | InterPro; IPR009638; Fez1.  |               |  |
| DR   | InterPro; IPR003900; KID repeat.  |               |  |
| DR   | InterPro; IPR001638; SBP_bac_3.   |               |  |
| DR   | Pfam; PF06818; Fez1; 1.   |               |  |
| DR   | Pfam; PF02524; KID; 5.  |               |  |
| DR   | PROSITE; PS01039; SBP_BACTERIAL_3; UNKNOWN_1.                           |               |  |
| KW   | Hypothetical protein.   |               |  |
| SQ   | SEQUENCE 4007 AA; 453246 MW; 67CDF6FEF1463612 CRC64;                    |               |  |
| Query Match 8.1%; Score 122.5; DB 2; Length 4007;                |   |               |  |
| Best Local Similarity 21.0%; Pred. No. 78;                       |   |               |  |
| Matches 74; Conservative 63; Mismatches 144; Indels 71; Gaps 14; |   |               |  |
| QY   | 8 QTVVVKSAIBTADGALDF-----YNKLDQVIPWKTDETIK-----ELSRFKQE 53              |               |  |
| DB   | 1156 QETELRKQHQSRVGESEIATIKYKKOLDLSRNTSQDAIKLKHENELANFKAK 1215          |               |  |
| QY   | 54 YQASVLVGDIKVLLMDSQDKYPEATQTVVCGVTVQLLSAVI-----LLFDYNEKKA 109         |               |  |
| DB   | 1216 YEQEKQLAVQHKTEMESLTDYHEKEKLATQYQERV-QALSALADKKYALAEYKE-QL 1273     |               |  |
| QY   | 110 SAQKDLIRILDGQVKNLNEAQKSL-----LGSSQSFNNASGKLALDSQ- 154               |               |  |
| DB   | 1274 SASAKQLDKLADHGKVDQLQAKLSEVAKVTADYEGNLSELRTKHQGEVNLKVHHD 1333       |               |  |
| QY   | 155 ----LTNDFSEKSYFQSQVDRIKEAYAGAAGIVAGPFGLIISYIAAGVIEGKL---- 207       |               |  |
| DB   | 1334 EIKKLTAGHNEKIRNLEHRLNDLKAELKQDRA-----EFDKKKALLEGVATL 1381          |               |  |
| QY   | 208 ----IPELNDRLKAVQNFTSL-----SVTVKQAKDIDAAKLKLTATEIAAIGEIKETET- 259    |               |  |
| DB   | 1382 QGKVDKSSKLSKSEKAEFNEKLKNEAQIAELRKVDKNSLQDKLEISDLKCOQKTR 1441       |               |  |
| QY   | 260 -TRFVVDYDDLMLSLLKG-----AAKMIN-TCNEYQORHG--KXTLEVPD 302              |               |  |
| DB   | 1442 IEDFNVQINERKMAQLLKAQNEIKASQASLNTTTTTEYDAKIAQLEKSLKEKKD 1493        |               |  |
| RESULT 17  |   |               |  |
| Q7RELO   | PRELIMINARY;  | PRT; 2664 AA. |  |
| AC   | Q7RELO;   |               |  |
| DT   | 01-MAR-2004 (Tremblrel. 26, Created)                                    |               |  |
| DT   | 01-MAR-2004 (Tremblrel. 26, Last sequence update)                       |               |  |
| DE   | Rhoptry protein (fragment).   |               |  |
| GN   | Name=PY05054;   |               |  |
| OS   | Plasmodium yoelii yoelii.   |               |  |



```
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteu M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DDJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AABL01001582; EAA17018.1; -.
DR InterPro: IPR006499; ReticulocyteP.
DR TIGRFAMs: TIGR01612; 235kDa-fam; 1.
FT NON TER 2664
SQ SEQUENCE 2664 AA; 312670 MW; FDE72DB05743F4AB CRC64;

Query Match      8.0%; Score 121; DB 2; Length 2664;
Best Local Similarity 21.3%; Pred. No. 60;
Matches 71; Conservative 67; Mismatches 140; Indels 56; Gaps 14;

QY 8 QTVEVVKSAIET--ADGALDFYNYKLDQVVPWK-TFDETIKELSRPK---QEYSQEAASVL 61
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1457 QYIATKEDLTHNDHSNISKNEHKDKSKAYKDEIDNNIKKLEENKKIPEKYQETTVL 1516

QY 62 VGDIKVLMSDQKFEATQTYVEGCVVTVLLSAY--ILLDFEYNEKKASAKQDILIRI 119
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1517 LNKYSALAKNK---FDQTKDSEL--IIKEIKDHKNCVVQSETSEQKMTNTIKNEQIFI 1571

QY 120 LDDGVN--KLNFAKSLGSSQSNFASGKLALDS--QLTNDFFSEKSSYFQSDVRLK 175
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1572 DSNNTNNDKSNNAIISIQSIEKFKT---KLLNNGIMQKNDCLKETTDIEEQISKU-- 1626

QY 176 BAYAGAAAGVAGPFGLIISYSIAAGVIEGKL-----IPELNDRLKAVQNPFSTLSV 227
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1627 -SINNQDTLVNNEIKLNDPKFKLSSLAKEKKIYIQDREKELNEVNSQIKNIEN----- 1678

QY 228 TVQANKDIDAKLKLATEIAAIGIKETETTRTFYVDYDMLSL-----L 274
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1679 DVNQYKDYEGIVEKINEIAKANKDKIESTKESIEPIQINIMSFNSDIEDIASNEAL 1738

QY 275 KGAAKMINTCNEYQORHG-----KKTLLVDPDI 303
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1739 ETYKTSMNNTNEFMESYDLIKKYSETVSKETPT 1772

RESULT 18
QY 7M918 PRELIMINARY; PRT; 693 AA.
ID Q7M918
AC Q7M918
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE PUTATIVE MEMBRANE PROTEIN.
GN OrderedLocusNames=WS0886;
OS Wolinella succinogenes.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Wolinella.
OX NCBI_TaxID=844;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSMZ 1740;

RX MEDLINE=22882897; PubMed=14500908;
RA Baar C., Eppinger M., Raddatz G., Simon J., Lanz C., Klimmek O.,
RA Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,
RA Meyer F., Lederer H., Schuster S.C.;
RT "Complete genome sequence and analysis of Wolinella succinogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695(2003).
DR EMBL: BX571659; CAE09993.1; -.
SQ SEQUENCE 693 AA; 77641 MW; CE57A1F23CFACCC2 CRC64;

Query Match      8.0%; Score 120.5; DB 2; Length 693;
Best Local Similarity 21.3%; Pred. No. 13;
Matches 64; Conservative 52; Mismatches 124; Indels 61; Gaps 9;

QY 11 EVVKSATETADGALDFYNYKLDQVVPWKTFDETIKELSRPKQEYSQEAASVLVGDIVLLM 70
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 EEVTSSTKEAQESLTQSSRWLQERIQ-----SDISLAQQHPTFHLGGLSTEFKALV 367

QY 71 DSQKYFEATQTYVEGCVVTVLLSAYILLDFEYNEKKASAKQDILIRIILDDGVNKLNEA 130
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 SLREQLEGANEVLSLLQKSSQESSAFLL----EQEKMAKESLTLVLFKAKKERFERLSEA 423

QY 131 -----OKSLGSSQSENN-----ASGKLLALDSLTNDFFSEKSSYFQSQV 170
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 FLTNKGALBEGNALDFRGFWEEYATRWRESSELQTGTLLETNQVRSFAELSEGVLQAN 483

QY 171 DRIRKEAYAGAAAGVAGPFG--GLIISYSIAAGVIEG--KLIPELNDRLKAVQNPFSTLSV 227
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 SAMGEQIKAG--PEGMTQALSGMLAKALEGGIDGVKKSVMENQTL----- 530

QY 228 TVQANKDIDAKLKLATEIAAIGIKETETTRTFYVDYDMLSLGAKAKKMTNCTNE 287
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 531 ----AKETLQTLQNHSSQSLGLLGESTTAIQTR-----LLEEBEAGKSLKLNLALE 577

QY 288 Y 288
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 578 Y 578

RESULT 19
QY 7VI137 PRELIMINARY; PRT; 555 AA.
ID Q7VI137
AC Q7VI137
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=HH0771;
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA Bell M., Droge M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J., Froesch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
RT Helicobacter hepaticus."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR EMBL: AE017146; AAP77368.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 555 AA; 61077 MW; 1E2515BCB83A804 CRC64;

Query Match      7.8%; Score 117.5; DB 2; Length 555;
Best Local Similarity 22.9%; Pred. No. 16;
Matches 83; Conservative 54; Mismatches 122; Indels 103; Gaps 18;

QY 13 VKSAIETADGALDFYNYKLDQV-----IPWKTFDETIKELSRPKQEYSQEAASVL 62
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 13 VKSAIETADGALDFYNYKLDQV-----IPWKTFDETIKELSRPKQEYSQEAASVL 62
```

```
Db 6 VKFTFNVKDSLQKIKKELDSINKNSANELYALVELKLHEQKMEHI-----QEYSKAYKDLQ 61
Qy 63 GDIVKLLMDS-----QKYPFATQVYEWGVVTPOLLISAVI--LLFDEYN 105
Db 62 SDIQSALNEDIQIPNGNVHYFRSLINELFSLQSVIT-KGFATSIISAAFWESSVIESMN 120
Qy 106 EKASQAQKILIRILDDGVNKLNEAQK-----SLLGSSQS---FNNASGKLLA--LDSQ 154
Db 121 KSLASAIQSLFSLNFD-KNRIANVQKEDWHLSEVVGAAFAFGIGNVAGGLVGNFLGDE 179
Qy 155 LTNDFSEKSSYFQSQVDRIKAEAVAGAAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNDR 214
Db 180 VNAQKTOKSA-----NNGAMAGAAAGAAIGSFVPIGTSL--GGLVGLVGLGGT 228
Qy 215 LKAVQNFFTLSVT-----VKQANKDIDAAKLKLATEIAAIGK-----TETETTR 261
Db 229 LFGSFN-STKLTTTAQGVLLISKATKDNVSAR-----EFADKEEVKKKMWGLQSNSTSWK 282
Qy 262 FYVD-----YDDL-----LSLKGAAKMINTCNEYQQRHGKTL 297
Db 283 EYDASNFAKGIQOOSIRGYEYLLQDIGSVKSLSTAKGNYSYADILNA-----GAKEL 337
Qy 298 LE 299
Db 338 IK 339

RESULT 20
O07569 PRELIMINARY; PRT; 2139 AA.
ID Q07569, O07504;
AC Q07569, O07504;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Myosin heavy chain.
DE Myosin heavy chain.
GN Name=hmhA;
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RX MEDLINE=93295430; PubMed=8515774;
RA Raymond-Denise A., Sansonetti P., Guillen N.;
RT Identification and characterization of a myosin heavy chain gene
RT (hmhA) from the human parasitic pathogen Entamoeba histolytica.";
RL Mol. Biochem. Parasitol. 59:123-131 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RA Guillen N.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; L03534; AAB48065.1; -.
DR PIR; T18296; T18296.
DR HSSP; P08799; 1MND.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00612; IQ; 2.
DR PRINTS; PR00193; MYOSINHEAVY.
DR SMART; SM00242; MYSC; 1.
SQ SEQUENCE 2139 AA; 245225 MW; C68307341DB51DD1 CRC64;

Query Match 7.7%; Score 117; DB 2; Length 2139;
Best Local Similarity 19.0%; Pred. No. 85;
Matches 75; Conservative 47; Mismatches 114; Indels 158; Gaps 13;

Qy 7 EQTVVVKSAIETADGALDFYKLDQVTPKTFDETIKELSRPKQYSEASVILGVGDIK 66
Db 965 EVEITELNSQINTLNATVD-----KDKTIAESQESIDEKEDEITKLGDIK 1011
```

```
Qy 67 VL-----LMSODKVFETATQ 81
Db 1012 LLEEKDDLEQRADVSATKDIAKKINKITIECEDAKDEIAKRLQELEDEENKKNKDLTN 1071
Qy 82 TVYEW---CGVTVQLLSAYILLFDEYNEKASQAQKILIRILDDGVNKLN-----E 129
Db 1072 ELQOTQLKLGTEKESLAAQVAA-----TKASDERDTLSQNLN--EKLTNNKLTKTAKD 1124
Qy 130 AQKSLIGSSQS-----NNASGKLLALDSOLTNDNFSEKSSYFQSQVDRIK 175
Db 1125 LEKISGLKQDVEDLEDDKNNKIEGDLRNAORKIKELDEITKG-ADVSQYLQK-----K 1178
Qy 176 EAVAGAAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNRLKAVQNFFTLSVTVQANKD 235
Db 1179 EYESQ-----IAKMQEKEAIGNDVKNKKEKTIKEKELE 1212
Qy 236 I-----DAAKLK--LATEIAAIGKIEKTETETTRFYVDYDLM----- 271
Db 1213 IQSLQEKLDTEVEKEDAEKKKEIKEMKALQOEKENVENSSKNSTKDKKKLEDNLKD 1272
Qy 272 -----SLLKGAAKMINTCNEYQQRHGK 294
Db 1273 QKLLDWTADNEKLIKAKAKDLEAQLNEVDNHEK 1306

RESULT 21
O064067 PRELIMINARY; PRT; 478 AA.
ID O64067, 064067;
AC O64067, 1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein yonD.
GN Name=yonD;
OS Bacteriophage SPBC2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=66797;
RN [1]
RP SEQUENCE FROM N.A.
RC Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Manuel C.,
RA Karamata D.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020713; AAC13027.1; -.
DR PIR; T12818; T12818.
KW Hypothetical protein.
SQ SEQUENCE 478 AA; 55050 MW; 6FF7495A957D4A4F CRC64;

Query Match 7.7%; Score 116.5; DB 2; Length 478;
Best Local Similarity 23.0%; Pred. No. 16;
Matches 58; Conservative 39; Mismatches 74; Indels 81; Gaps 12;

Qy 25 DFYKRY-LDQVTPKTFDETIKELSRPKQYSEASVILGVGDIKVLMSODKVFETATQV 83
Db 268 DVIDTYFIVNVVWS--DENSVD-KYFKFNTR-----TGDVTSIDFDSKTEVF-----MT 315
Qy 84 YEWCGVTVQLLSAYILLFDEYNEKASQAQKILIRILDDGVNKLNEAQKSLIGSSQSFPNN 143
Db 316 RNWEEVPEPIQSOL-----NQKDEIQKLTQVQNQIN---KDKVGIEQQENT 360
Qy 144 ASGKLLALDSQ-----LTNDFSEKSSYFQSQVDRIKAEYAGAAAGIVAG 188
Db 361 ASEKLVQLNSEVQQLPKPYKEKHEKTLLEQKLSEKNEFYKAKFEALNAEE----- 409
Qy 189 PFGLIISYSIAAGVIEGKLIPELNRLKAVQNFFTLSVTVQANKDIDAAKLKLATEIA 248
Db 410 -----KFS-----TBEVQNL-----IHASVKQ--DEEGEKAVQLNTMLV 442
Qy 249 AIGELKTETETT 260
Db 443 DIVSVPTETNTT 454
```

```

RESULT 22
ID O31954
AC O31954;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DN YonD protein.
GN Name=yonD; OrderedLocusNames=BSU21130;
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=980404033; PubMed=9384377; DOI=10.1038/36786;
RA Kunet F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borris R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerth I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Enlian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M.,
RA Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaerz-Blanchard M.,
RA Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,
RA Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,
RA Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C.,
RA Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,
RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,
RA Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,
RA Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G.,
RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,
RA Rose M., Sadate Y., Sato T., Scanlan E., Schleich S., Schroeter R.,
RA Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Seror P.,
RA Shin B.S., Soldo B., Sorokin A., Taccioni E., Takagi T., Takahashi H.,
RA Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P.,
RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,
RA Vasarotti A., Viari A., Wambutt R., Wedler E., Wedler H.,
RA Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K.,
RA Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E.,
RA Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).
DR EMBL; Z99115; CAB14031.1; -.
KW Complete proteome.
SQ SEQUENCE 478 AA; 55050 MW; 6FF7495A957D4A4F CRC64;

Query Match 7.7%; Score 116.5; DB 2; Length 478;
Best Local Similarity 23.0%; Pred. No. 16;
Matches 58; Conservative 39; Mismatches 74; Indels 81; Gaps 12;

Qy 25 DFYKY-LDQVTPKTFDETIELSRFKQVYSEASVLGDIKVLMDSDQKYFEATQV 83
Db 268 DVIDYFIVNYSWS--DENSD-KYFKFNTR-----TGDVTSIDFSKTEVF---MT 315

Qy 84 YEWCGVTTQLSAYILLFDEYNEKKASAKQDILIRILDGQVKNLNEAKSLGSSQSPNN 143
Db 316 RNWEEVPEPIQSOL-----NQKDEIKDLTKQVQNIN---KDKVGIEQFNT 360

Qy 144 ASGLLALDSQ-----LITNDFSEKSSYFQSDVRIRKEAYAGAAGIVAG 188
Db 361 ASEKLVLNLSVEQQLKPYKEHKHTLLEQKSEKNEFYKAKFEALNASE----- 409

Qy 189 PGLIISYIAAGVTEGKLIPELNDRLKAVQNFFTSLSVTKQANKDIDAAKLKLAIEIA 248
Db 410 -----KFS-----TEEVQNL-----IHASVKQ-DEGEKAVQLQNTMLV 442

Qy 249 AIGEIKTETTT 260

```

```

Db 443 DLVSVPTETNTT 454

RESULT 23
Q7CMF0 PRELIMINARY; PRT; 652 AA.
ID Q7CMF0
AC Q7CMF0;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE S-layer protein, (PX01-90).
GN Name=EXA0124;
OS Bacillus anthracis str. A2012.
OG Plasmid pX01.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=191218;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A2012;
RX MEDLINE=22061436; PubMed=12004073;
RA Read T.D., Salzberg S.L., Pop M., Shumway M., Umayam L., Jiang L.,
RA Holtzapfel E., Busch J.D., Smith K.L., Schupp J.M., Solomon D.,
RA Keim P., Fraser C.M.;
RA "Comparative Genome Sequencing for Discovery of Novel Polymorphisms in
RT Bacillus anthracis";
RL Science 296:2028-2033(2002).
DR EMBL; AE011190; AAM26077.1; -.
DR InterPro; IPR001119; SLH.
DR InterPro; IPR010989; t-snare.
DR InterPro; IPR010978; tRNA_binding_arm.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS01072; SLH_DOMAIN; UNKNOWN_1.
KW Plasmid.
SQ SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;

Query Match 7.7%; Score 116.5; DB 2; Length 652;
Best Local Similarity 19.0%; Pred. No. 22;
Matches 59; Conservative 65; Mismatches 106; Indels 81; Gaps 10;

Qy 25 DFYKYLDDVTPKTFDETIELSRFKQVYSEASVL-----VGDIKVLMDSDQKYFE 78
Db 237 DVINQKIDE---FDKLSQRKDLERMLEELNQLSKLQKQSPQLDNLKNKESQSRLL 293

Qy 79 ATQ-----TVYEWCGVTTQLSAYILLFDEYNE--KKASAKQDILIRILDGQVKN 126
Db 294 LNKXDSNRLELNSKIKKLNDRKAEILLSLIMELIKQSEFDKKIKNEKDDLKRELLNR 353

Qy 127 LNEAQK-----SLGSSQSPNNASGKLLALDSQLTNDFSEKSSYFQ---SQV 170
Db 354 IAESKELAKKKAELNTKLVELFKVQEAALNKKSQVLYYINKLDNELRELADKYKNSDKI 413

Qy 171 DRIRKEAYAGAAGIVAGPFGLIISYSIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTK 230
Db 414 SRLKNH-----IGYKNQKLEKLEN-----BLE 435

Qy 231 QANKDIDAAKLKLAT-EIAAIGEIKTETTTFRFVYVDL-----MLSLGAKAKMINT 284
Db 436 ECNKIDNTKKQALAEFDKSNKKQQLSESELVQLNKKIDELGKRHKHGRQLEASQKALDE 495

Qy 285 CNEYQQRHGKK 295
Db 496 AKINKKLAEK 506

RESULT 24
ID Q9X360
AC Q9X360;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DE 01-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE PX01-90 (S-layer protein,).

```

```

GN OrderedLocusNames=GBAA_PX01_0124;
OS Bacillus anthracis.
OC Plasmid virulence plasmid PX01, and plasmid pX01.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Strom, PLASMTD=virulence plasmid PX01;
RX MEDLINE=99445483; PubMed=10515943;
RA Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,
  Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Manter D.,
  Martinez Y., Ricke D., Svensson R., Jackson P.J.;
RT "Sequence and organization of pX01, the large Bacillus anthracis
  plasmid harboring the anthrax toxin genes.";
RL J. Bacteriol. 181:6509-6515(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate 0581; PLASMTD=pX01;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
  Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
  Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065404; AAD32394.1; -
DR EMBL; AE017336; AAT28865.2; -
DR PIR; B59102; B59102.
DR TIGR; GBAA_PX01_0124; -.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS01072; SLH DOMAIN; UNKNOWN_1.
DR Complete proteome; Plasmid.
KW Complete proteome; Plasmid.
SQ SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;

Query Match 7.7%; Score 116.5; DB 2; Length 652;
Best Local Similarity 19.0%; Pred. No. 22;
Matches 59; Conservative 65; Mismatches 106; Indels 81; Gaps 10;

QY 25 DFYKYLQVDPKWTFTDETIKELSRFKQEQYSQASVL-----VGDIKVLLMDSQDKYFE 78
DB 237 DVINKIDE---FDKLSQRKDLERMLELNQKLSQKQSPQLQDLNKNLKESSRLLE 293

QY 79 ATQ-----TVYEWGVVTVLLSAYILLFDEYNE--KKASAQKDLIRILDDGVNK 126
DB 294 LNKDLSNRLEINSEIKLNDRAKELLSLIMELIKQSEFDDKIKNEKDDLNKREDLINR 353

QY 127 LNEAQK-----SLGSSQSFNNASGKLLALDSQLTNDFSEKSSYFQ---SQV 170
DB 354 IAESKELAKKKAELNKLVELFKVQEAALNKKSGQYLYINKLDNELRELADKYKNSDNKI 413

QY 171 DRIRKAVAGAAAGIVAGPGLIISYIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTVK 230
DB 414 SRLKNH-----IGEYNKQLEKIE-----ELE 435

QY 231 QANKOIDAAKLLKAT-ETIAIGEIKTETTTTFYVDYDDL-----MLSLKGAACKMINT 284
DB 436 ECNKIDNTTKQLABFDKSNKKQOELESELVQLNKKIDELGKRHKHQRHLEASQKKALDE 495

QY 285 CNEYQQRHGKK 295
DB 496 AKEINKKLAEK 506

RESULT 26
Q7RM79 PRELIMINARY; PRT; 779 AA.
AC Q7RM79;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DE GTPase of unknown function, putative.
GN Name=PY02306;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Anguoli S.V., Suh B.B., Koolij T.W., Perteza M.,
  Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
  Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
  Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
  Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
  Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
  Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
  van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
  Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
  Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
  parasite plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an

```

```

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
  Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017336; AAT28865.2; -
KW Plasmid.
SQ SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;

Query Match 7.7%; Score 116.5; DB 2; Length 652;
Best Local Similarity 19.0%; Pred. No. 22;
Matches 59; Conservative 65; Mismatches 106; Indels 81; Gaps 10;

QY 25 DFYKYLQVDPKWTFTDETIKELSRFKQEQYSQASVL-----VGDIKVLLMDSQDKYFE 78
DB 237 DVINKIDE---FDKLSQRKDLERMLELNQKLSQKQSPQLQDLNKNLKESSRLLE 293

QY 79 ATQ-----TVYEWGVVTVLLSAYILLFDEYNE--KKASAQKDLIRILDDGVNK 126
DB 294 LNKDLSNRLEINSEIKLNDRAKELLSLIMELIKQSEFDDKIKNEKDDLNKREDLINR 353

QY 127 LNEAQK-----SLGSSQSFNNASGKLLALDSQLTNDFSEKSSYFQ---SQV 170
DB 354 IAESKELAKKKAELNKLVELFKVQEAALNKKSGQYLYINKLDNELRELADKYKNSDNKI 413

QY 171 DRIRKAVAGAAAGIVAGPGLIISYIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTVK 230
DB 414 SRLKNH-----IGEYNKQLEKIE-----ELE 435

QY 231 QANKOIDAAKLLKAT-ETIAIGEIKTETTTTFYVDYDDL-----MLSLKGAACKMINT 284
DB 436 ECNKIDNTTKQLABFDKSNKKQOELESELVQLNKKIDELGKRHKHQRHLEASQKKALDE 495

QY 285 CNEYQQRHGKK 295
DB 496 AKEINKKLAEK 506

RESULT 26
Q7RM79 PRELIMINARY; PRT; 779 AA.
AC Q7RM79;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DE GTPase of unknown function, putative.
GN Name=PY02306;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Anguoli S.V., Suh B.B., Koolij T.W., Perteza M.,
  Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
  Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
  Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
  Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
  Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
  Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
  van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
  Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
  Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
  parasite plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an

```

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.

CC EMBL; AABL01000631; EAA21743.1; --  
DR GO; GO:0005525; F:GTP binding; IEA.  
DR InterPro; IPR005289; GTP-binding.  
DR InterPro; IPR002917; MMR\_HSR1.  
DR Pfam; PF01926; MMR\_HSR1; 1.  
DR TIGRFAMs; TIGR00650; MG442; 1.  
SQ SEQUENCE 779 AA; 92336 MW; 7F51893C396B0B52 CRC64;

Query Match 7.7%; Score 116.5; DB 2; Length 779;  
Best Local Similarity 20.3%; Pred. No. 28;  
Matches 70; Conservative 51; Mismatches 126; Indels 97; Gaps 14;

QY 27 YNKYLQVDPKFTFDTIKELSRFKQYSQEAASVLVGDIKVLLMDSQDKYFEATQTVYEW 86  
DB 95 YGKEVENTEFFQTFDEYFN--SRSEKEDYEQ---LEAGKLI-LTENTYKDEIRKIME- 147  
QY 87 CGVVTQLLSAYILLDFEYNEKASAKO-----ILIRILD-----G 123  
DB 148 -----KSY---KSEYEKLKBEEDSKISHDVHENLKVTFVMKKNIDNPLFTQYSYM 196  
QY 124 VNKLNEAKSLGSSOSFNNSAKLLALDSQLTND-----FSEKSYF-----QSQVDRIR 174  
DB 197 INKLKSEILKSLKNVFNNGSTTSNIENNDQVCLDEKSEKYNDEKSKSLREIN 256  
QY 175 KEAYAGAAAGVAGPFGLIISYSIAAGVIEGKLIPELN-----DRLKAVQNFETSL 226  
DB 257 KPGDAENA-----LEPKYKSGNKENKLIBEVNFFPKQGYAEKRISTSDQMF 305  
QY 227 VTVKQANK-----DIDAAKLKATEIAAIGIKTETETRYVPYVDLMLSLKGA 277  
DB 306 ANVKIENKLLIKGEYDNDGGLKLEGEVVISYNNNEDEGTELFKNVNSVLELFKEN 365  
QY 278 AKMI-----NTCNE-----YQQRHGKTKILLEVPD 302  
DB 366 VKASLYKKENIKNNLNLGKRVKVVHFFPKMKRIIMKIPD 409

RESULT 27  
Q74DE2 PRELIMINARY; PRT; 541 AA.

ID Q74DE2  
AC Q74DE2  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Methyl-accepting chemotaxis protein.  
GN Name=hyLB; ORFNames=GSU1374;  
OS Geobacter sulfurreducens.  
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;  
OC Geobacteraceae; Geobacter.  
OX NCBI\_TaxID=35554;  
RX [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=PCA / ATCC 51573;  
RX PubMed=14671304; DOI=10.1126/science.1088727;  
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C., Heidelberg J.F., Wu D., Ward N.L., Beanan M.J., Dodson R.J., Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S., Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J., Daviden T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A., Weidman J.F., Khouri H.M., Feldblyum T.V., Utterback T.R., Van Aken S.E., Lovley D.R., Fraser C.M.;  
RT "Genome of Geobacter sulfurreducens: metal reduction in subsurface environments.";  
RL Science 302:1967-1969(2003).  
DR EMBL; AE017180; AAR34750.1; --  
DR TIGR; GSU1374; --  
DR InterPro; IPR004089; Chmtaxis\_transd.  
DR InterPro; IPR003660; HAMP.  
DR InterPro; IPR004090; We.chemotaxis.  
DR InterPro; IPR005829; Sug.transporter.  
DR InterPro; IPR010989; t-snar.

DR Pfam; PF00672; HAMP; 1.  
DR PRINTS; PF00015; MCPsignal; 1.  
DR SMART; PRO0260; CHEMTRNSDUCR.  
DR SMART; SM00304; HAMP; 1.  
DR SMART; SM00283; MA; 1.  
DR PROSITE; PS00111; CHEMOTAXIS\_TRANSDUC\_2; 1.  
DR PROSITE; PS00885; HAMP; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
SQ SEQUENCE 541 AA; 57093 MW; 3C9BEDAEAF7C3812 CRC64;

Query Match 7.7%; Score 116; DB 2; Length 541;  
Best Local Similarity 22.7%; Pred. No. 19;  
Matches 66; Conservative 42; Mismatches 111; Indels 72; Gaps 10;

QY 19 TADGALDFYNKYLDQVLPKFTFDTIKELSRFKQYSQEAASVLVGDIKVLLMDSQDKYFE 78  
DB 36 TANNGLD--TVYRDRVLP-----LKDLKIADMY-----AVNIVDVSHKVRNGNITWTE 82  
QY 79 ATQTVYEWCGVVTQLLSAYILLDFEYNEKASAKOILIRILDGWNKLNBAQKSLGSS 138  
DB 83 GRKSVEEAKKTIABKLOAYLATNLAAEEKKHLEAKPLIKVAD-----ATLERIASI 134  
QY 139 QSFNNASCKLLALDSQLTNDSEKSSYFQSOVD---RIRKEAY-----A 179  
DB 135 LSAEDAEALTFTVSELYPAIDPVSAKFSLSLVDQLKIAQDEYDHSSGLYRASRTISLVA 194  
QY 180 GAAAGVAGPFGLIISYSIAAGVIEGKLI-----PELNDRLKAVQNF 221  
DB 195 IIVGLVINGTAGLITRISITGFLAEGVEVANRLAAGDLTVEVRAGGRDETQOLMAAGNM 254  
QY 222 FTSL-----SVTVKQANKDIDAALKLAT---EIAA-IGEIKTETE 258  
DB 255 VTSRLHLIAEASISHGFIASASNLHATSEQIATGSEBVAQVGAVATASE 305

RESULT 28  
AAR34750 PRELIMINARY; PRT; 541 AA.

ID AAR34750  
AC AAR34750;  
DT 02-MAR-2004 (TREMBlrel. 27, Created)  
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)  
DE Methyl-accepting chemotaxis protein.  
GN HVLB OR GSU1374.  
OS Geobacter sulfurreducens.  
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;  
OC Geobacteraceae; Geobacter.  
OX NCBI\_TaxID=35554;  
RX [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=PCA / ATCC 51573;  
RX PubMed=14671304;  
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C., Heidelberg J.F., Wu D., Ward N.L., Beanan M.J., Dodson R.J., Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S., Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J., Daviden T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A., Weidman J.F., Khouri H.M., Feldblyum T.V., Utterback T.R., Van Aken S.E., Lovley D.R., Fraser C.M.;  
RT "Genome of Geobacter sulfurreducens: metal reduction in subsurface environments.";  
RL Science 302:1967-1969(2003).  
DR EMBL; AE017211; AAR34750.1; --  
DR TIGR; GSU1374; --  
DR InterPro; IPR004089; Chmtaxis\_transd.  
DR InterPro; IPR003660; HAMP.  
DR InterPro; IPR004090; We.chemotaxis.  
DR InterPro; IPR005829; Sug.transporter.  
DR InterPro; IPR010989; t-snar.

```
QY 79 ATQTVYVWGVVTTQLLSAYILLDFEYNEKKASAKDILIRLLDDGVNKLNEAKSLLGSS 138
Db 83 GKSVVEEAKKTAIEKLOAYLATNLAEEKGLLEEAKPLIKVAD-----ATLERLASI 134
QY 139 OSFNNASGKLALDSQLTNDSEKSSYFQSQVD---RIRKEAY-----A 179
Db 135 LSAEDAEALTFTVSELVPAIDPVSAKPSLVDOLKIAQEHSSGLYRASRTISLVA 194
QY 180 GAAAGIVAGPGLIISYSIAAGVIEGKLI-----PELNDRLKAVONF 221
Db 195 IIVGVLIAGTAGLITRISITGLAEVGEVANRLAAGDLTVFVAGGRDETQOLMAAGNM 254
QY 222 FTSL-----SVTVKQANKDIDAAKLKLAT---BIAA-IGRIKETE 258
Db 255 VTSRLHLTAELAISISHGIASASNLHATSEQIATGSEVSAQVGAVATASE 305

RESULT 29
Q81298
ID Q81298 PRELIMINARY; PRT; 941 AA.
AC Q81298, Q61440; G6KXV0;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2004 (TREMELrel. 28, Last annotation update)
DE Conserved domain protein.
CN OrderedLocNames=BA0374, BAS0360; ORFNAMES=GBAA0374;
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate Porton;
RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;
RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapfel E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,
RA DeBoy R.T., Madpu R.J., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,
RA Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M.,
RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria."
RL Nature 423:81-86(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA Fraser C.M.;
RT "Bacillus anthracis comparative genomics."
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Stemne;
RA Bretin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017025; AAP24404.1; -
DR EMBL; AB017334; AAT29469.1; -
DR EMBL; AB017225; AAT52691.1; -
DR TIGR; BA0374; -
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR008994; Nucleic acid OB.
DR SEQUENCE 941 AA; 103896 MW; 78AF6B5381D7B6C6 CRC64;

Query Match 7.7%; Score 116; DB 2; Length 941;
Best Local Similarity 20.1%; Pred. No. 37;
```

```
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7;
QY 7 EQTVVVKSAIETADGAL-DFYNKYLDQVIP-WKTFDETIKELSRFKQYSQEASVLVGD 64
Db 596 KEVVESINEATQNASAQGLDFLATYDSEIIVNFNTAERTKMSKNTSQILKEADKKLPD 655
QY 65 IKVLLMDSQDKYFEATQTVYVWGVVTTQLLSAYILLDFEYNEKKASAKDILIRLLDDGV 124
Db 656 VKLLEDSK-----GLV-----DGR 671
QY 125 NKLENAQSKLLGSSQSFNNASGKLALDSQ-----LTNDFSEKSSYFQSQVDRIK 175
Db 672 KKLADIKAEMPATEKKIKELADKIRDFESEBDLKDIIIRLLKNDVEKQSDYFANPNLKEN 731
QY 176 EAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELND---RLKAVQNF---TSLSVT 228
Db 732 KLFAMPNYGSA MSPFTVTLALWVGLLWVSLITVEVHEEGANYKSHIYFGRLLTFLTMG 791
QY 229 VKQA 232
Db 792 LSQA 795

RESULT 30
AAT29469
ID AAT29469 PRELIMINARY; PRT; 941 AA.
AC AAT29469;
DT 01-JUN-2004 (TREMELrel. 27, Created)
DT 01-JUN-2004 (TREMELrel. 27, Last sequence update)
DT 01-JUN-2004 (TREMELrel. 27, Last annotation update)
DE Conserved domain protein.
CN GBAA0374.
OS Bacillus anthracis str. Ames 0581.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics."
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017334; AAT29469.1; -
DR SEQUENCE 941 AA; 103896 MW; 78AF6B5381D7B6C6 CRC64;

Query Match 7.7%; Score 116; DB 2; Length 941;
Best Local Similarity 20.1%; Pred. No. 37;
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7;
QY 7 EQTVVVKSAIETADGAL-DFYNKYLDQVIP-WKTFDETIKELSRFKQYSQEASVLVGD 64
Db 596 KEVVESINEATQNASAQGLDFLATYDSEIIVNFNTAERTKMSKNTSQILKEADKKLPD 655
QY 65 IKVLLMDSQDKYFEATQTVYVWGVVTTQLLSAYILLDFEYNEKKASAKDILIRLLDDGV 124
Db 656 VKLLEDSK-----GLV-----DGR 671
QY 125 NKLENAQSKLLGSSQSFNNASGKLALDSQ-----LTNDFSEKSSYFQSQVDRIK 175
Db 672 KKLADIKAEMPATEKKIKELADKIRDFESEBDLKDIIIRLLKNDVEKQSDYFANPNLKEN 731
QY 176 EAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELND---RLKAVQNF---TSLSVT 228
Db 732 KLFAMPNYGSA MSPFTVTLALWVGLLWVSLITVEVHEEGANYKSHIYFGRLLTFLTMG 791
QY 229 VKQA 232
Db 792 LSQA 795

RESULT 31
HLV1_ECOLI
```

HLX1\_ECOLI STANDARD; PRT; 1023 AA.  
 P09983;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Hemolysin, chromosomal.  
 GN Name=hlyA;  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]\_TaxID=562;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J96 / Serotype O4;  
 RX MEDLINE=85234404; PubMed=3891743;  
 RA Felmlee T., Pellett S., Welch R.A.;  
 RT "Nucleotide sequence of an Escherichia coli chromosomal hemolysin.";  
 RL J. Bacteriol. 163:94-105(1985).  
 RN [2]  
 RP SEQUENCE OF 1-44 FROM N.A.  
 RC STRAIN=2001;  
 RX MEDLINE=85258115; PubMed=3894051;  
 RA Nicaud J.-M., Mackman N., Gray L., Holland I.B.;  
 RT "Characterisation of HlyC and mechanism of activation and secretion of  
 haemolysin from E. coli 2001.";  
 RL FEBS Lett. 187:339-344(1985).  
 CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood  
 cell membranes and cause cell rupture by mechanisms not clearly  
 defined.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- DOMAIN: The Gly-rich region is probably involved in binding  
 calcium, which is required for target cell-binding or cytolytic  
 activity.  
 CC -!- DOMAIN: The three transmembrane domains are believed to be  
 involved in pore formation by the cytotoxin.  
 CC -!- PTM: Palmitoylated by hlyC. The toxin only becomes active when  
 modified.  
 CC -!- MISCELLANEOUS: The hemolysin of E.coli is produced predominantly  
 by strains causing extraintestinal infections, such as those of  
 the urinary tract.  
 CC -!- SIMILARITY: Belongs to the RTX prokaryotic toxin family.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M10133; AAA23975.1; -;  
 CC EMBL; X02768; CAA26546.1; -;  
 CC PIR; A24433; LEECA.  
 CC InterPro; IPR001343; Hemlysn\_Ca\_bind.  
 CC InterPro; IPR003995; RTXA.  
 CC InterPro; IPR011049; Serralysin\_like\_C.  
 CC Pfam; PF00353; HemolysinCbind; 6.  
 CC Pfam; PF02382; RTX; 1.  
 CC PRINTS; PR00313; CAENDNGRPT.  
 CC PRINTS; PR01488; RTXTOXINA.  
 CC PROSITE; PS00330; HEMOLYSIN\_CALCIDIUM; 4.  
 CC Calcium; Cytolysis; Hemolysis; Lipoprotein; Palmitate; Repeat; Toxin;  
 KW Transmembrane.  
 FT TRANSMEM 237 259 Potential.  
 FT TRANSMEM 267 326 Potential.  
 FT TRANSMEM 364 410 Potential.  
 FT DOMAIN 723 869 16 X REPEATS, GLY-RICH.  
 FT REPEAT 723 728 1.  
 FT REPEAT 732 737 2.  
 FT REPEAT 741 746 3.  
 FT REPEAT 750 755 4.  
 FT REPEAT 759 764 5.  
 FT REPEAT 768 773 6.

FT REPEAT 777 782 7.  
 FT REPEAT 786 791 8.  
 FT REPEAT 795 800 9.  
 FT REPEAT 806 812 10.  
 FT REPEAT 816 821 11.  
 FT REPEAT 825 830 12.  
 FT REPEAT 834 839 13.  
 FT REPEAT 843 848 14.  
 FT REPEAT 855 860 15.  
 FT REPEAT 864 869 16.  
 FT LIPID 563 563 N(6)-palmitoyl lysine (By similarity).  
 FT LIPID 589 589 N(6)-palmitoyl lysine (By similarity).  
 FT VARIANT 6 A -> T (in strain 2001).  
 SQ SEQUENCE 1023 AA; 109867 MW; 196D5C0A9A28B54D CRC64;  
 Query Match 7.7%; Score 116; DB 1; Length 1023;  
 Best Local Similarity 20.2%; Pred. No. 41;  
 Matches 68; Conservative 59; Mismatches 112; Indels 98; Gaps 11;  
 QY 2 TGIFAEQTVVVK-----SAIETADGALDFVYKYLQVIPWKTFTETIKELSRFKQE 53  
 DB 155 TALSSMKIDELIKKQKSGNVSSSELAKASTELINQLVDTA-----ASLNNVNSFSQ 207  
 QY 54 YSOEASVL-----VGDIKVLLMDSQDKYFEATQTVYEWGVVTVLLSAYILL-FDEY 104  
 DB 208 LNKGLSVLSNTKHLNGVGN-KLQNLPLNDTIGAGLDTV---SGILSAISASFILSNADAD 263  
 QY 105 NEKSAQOKDILIRILDO---GVNKLNEAQKSLGSSOSFNNASGKLLALDLSQLTNDFSE 161  
 DB 264 TGTRAAAGVEITTKVLGVNGKISQYIIAQRAAQGLST----- 302  
 QY 162 KSYFQSQVDRIKREAYAGAAGIVAGPFGLIIS----YSTAAGVIEGKLIPELNDRLKA 217  
 DB 303 -----AAAAGLIASVVTLLAISPLSLFIADKPKRANKIEYSQRPKK 344  
 QY 218 VQNFPTSLSVTVKQANKDIDAQKLATEIAAIGIKETETETTRFYVYDDMLSLKGA 277  
 DB 345 LGYDGSLSLAAAFHKETGAIDASLTRISTVLASVSGISAAATTSLV----GAPVSALVGA 400  
 QY 278 -----AKQMINTCNEYQOORHGK 294  
 DB 401 VTGIISGILEASKQAMPEHVASKADVAEWEKKGK 437  
 RESULT 32  
 ID\_RASO\_METJA STANDARD; PRT; 1005 AA.  
 AC Q58718;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE DNA double-strand break repair rad50 ATPase.  
 GN Name=rad50; OrderedLocNames=MJ1322;  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=868087;  
 RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagen N.S.W., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RT Science 273:1058-1073(1996).  
 RL -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The



| ID                    | Q928G7   | PRELIMINARY; | PRT; 1093 AA. |
|-----------------------|--|--------------|---------------|
| AC                    | 01-DEC-2001 (TREMBlrel. 19, Created)                                 |              |               |
| DT                    | 01-DEC-2001 (TREMBlrel. 19, Last sequence update)                    |              |               |
| DT                    | 01-WAR-2004 (TREMBlrel. 26, Last annotation update)                  |              |               |
| DE                    | Lin2568 protein.   |              |               |
| GN                    | OrderedLocusNames=lin2568;   |              |               |
| OS                    | Listeria innocua.  |              |               |
| OC                    | Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.            |              |               |
| NCBI_TaxID=1642;      |  |              |               |
| RN                    | [1]  |              |               |
| RP                    | SEQUENCE FROM N.A.   |              |               |
| RC                    | STRAIN=CLIP 11262 / Serovar 6a;                                      |              |               |
| RX                    | MEDLINE=21537279; PubMed=11679669;                                   |              |               |
| RA                    | Glaser P., Prangeul L., Buchrieser C., Rusniok C., Amend A.,         |              |               |
| RA                    | Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,        |              |               |
| RA                    | Charbit A., Chetoui F., Couve E., de Daruvar A., Dehieux P.,         |              |               |
| RA                    | Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O., |              |               |
| RA                    | Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,          |              |               |
| RA                    | Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D., |              |               |
| RA                    | Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,     |              |               |
| RA                    | Madueno E., Maitournan A., Mata Vicente J., Ng E., Medjari H.,       |              |               |
| RA                    | Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  |              |               |
| RA                    | Remmel B., Rose M., Schluerter T., Simoes N., Tierrez A.,            |              |               |
| RA                    | Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;               |              |               |
| RT                    | "Comparative genomics of <i>Listeria</i> species."                   |              |               |
| RL                    | Science 294:849-852 (2001).  |              |               |
| DR                    | EMBL: AL596172; CAC97795.1; .  |              |               |
| DR                    | PIR: A1753; A1753.   |              |               |
| DR                    | Listinist; LIN2568; .  |              |               |
| DR                    | InterPro; IPR010090; Tape_meas_TP901.                                |              |               |
| DR                    | InterPro; IPR007713; TMP.  |              |               |
| DR                    | Pfam; PF05017; TMP; 11.  |              |               |
| DR                    | TIGRfams; TIGR01760; tape_meas_TP901; 1.                             |              |               |
| KW                    | Complete proteome.   |              |               |
| SQ                    | SEQUENCE 1093 AA; 121152 MW; 67AAA5905889B4D4 CRC64;                 |              |               |
| Query Match           | 7.6%; Score 115; DB 2; Length 1093;                                  |              |               |
| Best Local Similarity | 22.6%; Pred. No. 52;   |              |               |
| Matches               | 71; Conservative   |              |               |
| Qy                    | 2 TGFAEQTVVWVKSIAETAGADLDFYKNYL-----                                 |              |               |
| Db                    | 270 TGRFNKNFGLVGRKAKTKDTIVDFYNTGLVDSVEEAGEALTQTRQLNDLNSDLESVT 329    |              |               |
| Qy                    | 32 DOVIPW-KTFD-----ETIKELSRFKQBYOEASVLVGDIKVLLMDSQKYFEATQTVVW 86     |              |               |
| Db                    | 330 EKALAFSKTFDSDMNETLRGANALMETYGLSAEQSFDLMTVGAQNGLNKTDELGDNLAEY 389 |              |               |
| Qy                    | 87 CGVVTQLLSAYILLFD-----EYN-EKASQAQDKILLRIILDDGVNKLNEAQSLLG 136      |              |               |
| Db                    | 390 SGQPKQNGYSQAQGFETLEAGLDSGANLKDNDLVKFEFGIRVSDSSEKAVEG---LG 445    |              |               |
| Qy                    | 137 SSQSFNNASGKLLALDSOLTNDSEKSYFQSDVRIRKEAYAGAAAGIVAGPGLI--- 193     |              |               |
| Db                    | 446 -----GKWTMYADMKDGYDNNELFANLATEINKVGEQERASIVSAIFGSLGED 496        |              |               |
| Qy                    | 194 --ISYSTAAGVIEGKLPELNDRLKAVONFTSL---SVTVQAKNDIAAKL--KLATE 246     |              |               |
| Db                    | 497 NTVKVLTAMG-----DLNGELGAVQGYKDDVKGASEKLTETNSKQDLTKWHELOTA 548     |              |               |
| Qy                    | 247 IAAIGEIKTETTT 260  |              |               |
| Db                    | 549 LAPIGEYLLLEANT 562   |              |               |
| RESULT 34             |  |              |               |
| Q08581                | PRELIMINARY;   |              |               |
| ID                    | Q08581   | PRT; 821 AA. |               |
| AC                    | 01-NOV-1996 (TREMBlrel. 01, Created)                                 |              |               |
| DT                    | 01-NOV-1996 (TREMBlrel. 01, Last sequence update)                    |              |               |
| DT                    | 01-OCT-2003 (TREMBlrel. 25, Last annotation update)                  |              |               |





```
DR PROSITE: PS00422; GRANINS_1; UNKNOWN_1.
KW Hypothetical protein. 176615 MW; 9AE7AB0AABF6EAE CRC64;
SQ SEQUENCE 1478 AA; 7.5%; Score 113.5; DB 2; Length 1478;
Query Match
Matches 68; Conservative 51; Mismatches 98; Indels 109; Gaps 15;
15;
QY 5 FAQTVVVKGAIEADGALDPYNYK-----DQVWPKTDFE-TIKELSRFKQ 52
DQVWPKTDFE-TIKELSRFKQ 52
Db 328 FFEKTLTEKEKEEYRNELDNKLLHAEQLNKKKLDQLERYKNDDEHIKSL----- 382
FFEKTLTEKEKEEYRNELDNKLLHAEQLNKKKLDQLERYKNDDEHIKSL----- 382
QY 53 EYSQASVLVGDIKVLLMDSQDKYFEATQTVVWGVVTVQLLSAVILLDFSYNEKKAQAQ 112
EYSQASVLVGDIKVLLMDSQDKYFEATQTVVWGVVTVQLLSAVILLDFSYNEKKAQAQ 112
Db 383 ---KESEIEINEKNTLILQQLAQAS---YE-----ISMENKSNKSN--- 422
KESEIEINEKNTLILQQLAQAS---YE-----ISMENKSNKSN--- 422
QY 113 KDILIRIILDDGVN---KLNKAEQSLGSSQSFNNASGKLLALDSQUTNDFSEKSYFQSQV 170
KDILIRIILDDGVN---KLNKAEQSLGSSQSFNNASGKLLALDSQUTNDFSEKSYFQSQV 170
Db 423 -----NNSYNGERNSEYEKKL---EELNN-----ITNSYEKEINELNKEK 459
NNSYNGERNSEYEKKL---EELNN-----ITNSYEKEINELNKEK 459
QY 171 DRIRK---EAVAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELND-----RLKAVQNF 221
DRIRK---EAVAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELND-----RLKAVQNF 221
Db 460 ENITRQFMKIMDDDEGI-----NKLKEELNDANALIVNLKNKNE 500
ENITRQFMKIMDDDEGI-----NKLKEELNDANALIVNLKNKNE 500
QY 222 FTSLSVTVKQANKDI-DAAKLKLAIEAIGRIKTKETETTRFYVDYDDLMLSLKGAAXK 280
FTSLSVTVKQANKDI-DAAKLKLAIEAIGRIKTKETETTRFYVDYDDLMLSLKGAAXK 280
Db 501 LYNVNIEMEOANKWRDDIDILLAN-----IDKLNDEKVTENEKEQKELKYNELKINYE 556
LYNVNIEMEOANKWRDDIDILLAN-----IDKLNDEKVTENEKEQKELKYNELKINYE 556
QY 281 MINTCN-----EYQQRH 292
MINTCN-----EYQQRH 292
Db 557 KYKECNKFFNPLPKMKKKIEYEKKH 582
KYKECNKFFNPLPKMKKKIEYEKKH 582

RESULT 37
RECN AQUAE STANDARD; PRT; 520 AA.
AC O66834;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA repair protein recN (Recombination protein N).
GN Name=recN; OrderedLocusNames=AQ_561;
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98196666; PubMed=9537320; DOI=10.1038/32831;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olsen G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -!- FUNCTION: May be involved in recombinational repair of damaged DNA
CC (By similarity).
CC -!- SIMILARITY: Belongs to the recN family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; A000695; AAC06789.1; -.
DR PIR; F70350; F70350.
DR HSSP; P25034; 1DEB.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR004604; RecN.
DR InterPro; IPR003405; SMC_C.
```

```
DR Pfam; PF02483; SMC_C; 1.
DR TIGRFAMs; TIGR00634; recN; 1.
KW ATP-binding; Complete proteome; DNA repair.
FT NP_BIND 29 36 ATP (Potential).
SQ SEQUENCE 520 AA; 60439 MW; A9708562ACBE901E CRC64;
Query Match
Matches 59; Conservative 70; Mismatches 94; Indels 71; Gaps 13;
13;
QY 10 VEVVKSIAETADGALDPYNYKLDQVWPKTDFE-----TIKELSRFKQYSQ---EASVL- 61
VEVVKSIAETADGALDPYNYKLDQVWPKTDFE-----TIKELSRFKQYSQ---EASVL- 61
Db 135 LEKVYSLRKKKEQLFELRKKKEELIKQDYLFERVREVEEIGISSEYEELKNKANLIN 194
LEKVYSLRKKKEQLFELRKKKEELIKQDYLFERVREVEEIGISSEYEELKNKANLIN 194
QY 62 -VGDIKVLMLDSQDKYFEATQTVVWGVVTVQLLSAVILLDFSYNEKKAQAQDI-- 115
VGDIKVLMLDSQDKYFEATQTVVWGVVTVQLLSAVILLDFSYNEKKAQAQDI-- 115
Db 195 NLEKVKKAVGSLYKLEGENSVYEIIGETIRKNLAKVESYSGKFSLEIKIANLEEEVYE 254
NLEKVKKAVGSLYKLEGENSVYEIIGETIRKNLAKVESYSGKFSLEIKIANLEEEVYE 254
QY 116 -----LIRILDDGVNKLNEAOKSLGSSQSFNNASGKLL-----ALDQSLTN---DF 159
LIRILDDGVNKLNEAOKSLGSSQSFNNASGKLL-----ALDQSLTN---DF 159
Db 255 LYNLSKEEMPEISEEVEINEKLFRIQRLKEKYKSPPEILKEVEEIKELSINLVDF 314
LYNLSKEEMPEISEEVEINEKLFRIQRLKEKYKSPPEILKEVEEIKELSINLVDF 314
QY 160 SEKSSYFQSQVDRIKRAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELN-DRLKAV 218
SEKSSYFQSQVDRIKRAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELN-DRLKAV 218
Db 315 KEEE--LREEVEKLREE-----YDKLAEEVSRRDKKA 345
KEEE--LREEVEKLREE-----YDKLAEEVSRRDKKA 345
QY 219 QNFFTSLSVTVKQANKDIDAOKLKLAIEAIGRIKTKETETTRFYVDYDDLMLSL 272
QNFFTSLSVTVKQANKDIDAOKLKLAIEAIGRIKTKETETTRFYVDYDDLMLSL 272
Db 346 EDLEERIEILKEUN--LERAKLV-----EIK-ESEPTKYGKDKIEFLFS 388
EDLEERIEILKEUN--LERAKLV-----EIK-ESEPTKYGKDKIEFLFS 388

RESULT 38
SVR SYNY3 STANDARD; PRT; 584 AA.
AC Q55486;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
GN Name=argS; OrderedLocusNames=all0502;
OS Synechocystis sp. (strain PCC 6803);
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugliura M., Tabata S.;
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; D64006; BAA10833.1; -.
DR PIR; S75986; S75986.
DR HSSP; Q93RP5; 1IQO.
DR HAMAP; MF_00123; -.
DR InterPro; IPR005148; ArgtrNASynthet_N.
DR InterPro; IPR001278; Arg tRNA-synt 1c.
DR InterPro; IPR008909; tRNA-synt_1d_C.
```

DR InterPro; IPR001412; tRNA-synt 1.  
 DR InterPro; IPR009080; tRNA-synt 1a bind.  
 DR Pfam; PF03485; Arg tRNA synt N; 1.  
 DR Pfam; PF00750; tRNA-synt 1d; 1.  
 DR Pfam; PF05746; tRNA-synt 1d C; 1.  
 DR PRINTS; PR01038; tRNA-synt 1d C; 1.  
 DR TIGRfam; TIGR00456; argS; 1.  
 DR PROSITE; PS00178; AA tRNA LIGASE I; 1.  
 KW Aminoacyl-tRNA synthetase; ATP-binding; Complete proteome; Ligase;  
 KW Protein biosynthesis.  
 FT SITE 126 136 "HIGH" region.  
 SQ SEQUENCE 584 AA; 65212 MW; 89602A1A5A3BD85B CRC64;  
 Query Match 7.5%; Score 112.5; DB 1; Length 584;  
 Best Local Similarity 23.2%; Pred. No. 36;  
 Matches 76; Conservative 45; Mismatches 131; Indels 75; Gaps 16;  
 QY 3 GIFAQTEVVVKSATETADGALD-----FY---NKLYDQVTPWTFDETIKELSRFKQE 53  
 DB 171 GMLITLKEVPEALVTAD-ALDIDGLVTFYKQAKQRFQD-----DEQFRETSR----- 218  
 QY 54 YSQEASVLVGDIK-----VLLMDSQDKYFEATQTVYEWGVVQ-----LLSAYIL 99  
 DB 219 -QAVVALQAGDAKSIKAWQLCEQSRREF---QLIYDCLDITIERGSEFYNPFLPGVVE 274  
 QY 100 LFDEYNEKKAQKDILIR-----ILDDGVNKLNEAQSLLGSSQ-SFNNAQKLLAL 151  
 DB 275 LLQE-----KOLLVEDNQAQCFLDGTNKGDRPLIVQKSDGGYNVATTDLAAL 325  
 QY 152 DSQLTNDPSEKSYF--QSQVDRIRKEAVAGAAAGIVAGPFLGLIISYIAAGVIEGKLIP 209  
 DB 326 NYRLNTDGAKEIYYVDAGQANHFQFFQVAKAGILTDPTQV---HVPFLVKGEDGK 382  
 QY 210 ELNDRLKAVQNFPTSLSVTVKQAKDID-----AAKLKATRIA---AIGELK---T 255  
 DB 393 KLKTRAGDTIRLKKLTTEAVFARQDLETRLTAEERSETEEPKTEVAQVGVGAVKYADL 442  
 QY 256 ETETRFYVDYDDLMLSLKGAKKMI 282  
 DB 443 SQNRTSDYVFPDKMLAQNTAPYML 469  
 RESULT 39  
 Q6CGC3 PRELIMINARY; PRT; 941 AA.  
 ID Q6CGC3  
 AC Q6CGC3  
 DT 01-OCT-2004 (TREMELrel. 28, Created)  
 DT 01-OCT-2004 (TREMELrel. 28, Last sequence update)  
 DE Similar to tr|Q8WZS2 Neurospora crassa Dynactin.  
 GN ORFNames=VALIOA20504g;  
 OS Yarrowia lipolytica (Candida lipolytica).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; Yarrowia.  
 OX NCBI\_TaxID=4952;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIB99;  
 RG GENOLEVRES;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E., Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C., Boisrame A., Boyer J., Catillico C., Confaniolieri F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dunazet H., Groppi A., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H., Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A., Swennens D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J., Wincker P., Souciet J.L.;  
 "Genome evolution in yeasts.";  
 RL Nature 430:35-44(2004).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RA STRAIN=CLIB99;  
 RA Genoscope;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CR382127; CAG84227.1;  
 SQ SEQUENCE 941 AA; 106891 MW; EF36A7BBE6D40AFB CRC64;  
 Query Match 7.5%; Score 112.5; DB 2; Length 941;  
 Best Local Similarity 23.4%; Pred. No. 84;  
 Matches 78; Conservative 49; Mismatches 118; Indels 89; Gaps 15;  
 QY 6 AEQTEVVVKSATET-ADGALDFYINKYLDQVFPW-----KTFDETIKELSRFKQESQAS 59  
 DB 547 SEQLEQSVNSYIDTVSEGAL-----LVDEAVNVPVFQTFDDFLSTQTDYNTYLDMSK 601  
 QY 60 VLVGD-IVKILM-----DSQDK-----YFEATQTVYEWGVV 91  
 DB 602 LAIHTRVSIILLKQLGHRSCANSLSKLYWIDNQDKTARVPSPITDNVQTLSDVSVPAM 661  
 QY 92 QLL-----SAVILLFDEYNEKKA-OKDILIRLLDDGVNKLNEAOKSLGSSQSFNA 144  
 DB 662 QLLHDDDPDSAYTAIFKRSSKRPJSAVEKDI-----KLTE-QLSTLDFSVSQND 711  
 QY 145 SGKLLALDSQLTNDPSEKSYFQSQVDRIRKEAVAGAAAGIVAGPFLGLIISYIAAGVIE 204  
 DB 712 ---LPTWAQLQPROSQEESIVDTSDVHLKAVL-----LKKETIETLEIK 755  
 QY 205 GKLIPELNDRLKAVQNFPTSL-----SVTVKQAKDIDAAKLKATEIAAIGIKET 257  
 DB 756 IKHYDSKAERFAEKEEAFVQLRANLNEAALTEKRLDDVESLKAKLEQKNADQORFKTHV 815  
 QY 258 ETRFYVDYDDLMLSLKLG--AAKGMINTCEYQ 289  
 DB 816 Q-----NRKLSTLKGEDIVDMHLNTHQSQ 841  
 RESULT 40  
 HLYA\_ECOLI  
 ID HLYA\_ECOLI STANDARD; PRT; 1024 AA.  
 AC P08715;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Hemolysin, plasmid.  
 GN Name=hlyA;  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hesse J., Wells W., Vogel M., Goebel W.;  
 RT "Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparison with a corresponding chromosomal hemolysin sequence.";  
 RL PEMS Microbiol. Lett. 34:1-11(1986).  
 RN [2]  
 RP PALMITOYLATION OF LYS-564 AND LYS-690.  
 RX MEDLINE=9509325; PubMed=7801126;  
 RA Stanley P., Packman L.C., Koronakis V., Hughes C.;  
 RT "Fatty acylation of two internal lysine residues required for the toxic activity of Escherichia coli hemolysin.";  
 RL Science 266:1992-1996(1994).  
 RN [3]  
 RP PALMITOYLATION OF LYS-564 AND LYS-690.  
 RX MEDLINE=96404790; PubMed=8808931;  
 RA Ludwig A., Garcia F., Bauer S., Jarchau T., Benz R., Hoppe J., Goebel W.;  
 RT "Analysis of the in vivo activation of hemolysin (HlyA) from Escherichia coli.";

RT "Genome evolution in yeasts.";  
 RL Nature 430:35-44(2004).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RA STRAIN=CLIB99;  
 RA Genoscope;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CR382127; CAG84227.1;  
 SQ SEQUENCE 941 AA; 106891 MW; EF36A7BBE6D40AFB CRC64;  
 Query Match 7.5%; Score 112.5; DB 2; Length 941;  
 Best Local Similarity 23.4%; Pred. No. 84;  
 Matches 78; Conservative 49; Mismatches 118; Indels 89; Gaps 15;  
 QY 6 AEQTEVVVKSATET-ADGALDFYINKYLDQVFPW-----KTFDETIKELSRFKQESQAS 59  
 DB 547 SEQLEQSVNSYIDTVSEGAL-----LVDEAVNVPVFQTFDDFLSTQTDYNTYLDMSK 601  
 QY 60 VLVGD-IVKILM-----DSQDK-----YFEATQTVYEWGVV 91  
 DB 602 LAIHTRVSIILLKQLGHRSCANSLSKLYWIDNQDKTARVPSPITDNVQTLSDVSVPAM 661  
 QY 92 QLL-----SAVILLFDEYNEKKA-OKDILIRLLDDGVNKLNEAOKSLGSSQSFNA 144  
 DB 662 QLLHDDDPDSAYTAIFKRSSKRPJSAVEKDI-----KLTE-QLSTLDFSVSQND 711  
 QY 145 SGKLLALDSQLTNDPSEKSYFQSQVDRIRKEAVAGAAAGIVAGPFLGLIISYIAAGVIE 204  
 DB 712 ---LPTWAQLQPROSQEESIVDTSDVHLKAVL-----LKKETIETLEIK 755  
 QY 205 GKLIPELNDRLKAVQNFPTSL-----SVTVKQAKDIDAAKLKATEIAAIGIKET 257  
 DB 756 IKHYDSKAERFAEKEEAFVQLRANLNEAALTEKRLDDVESLKAKLEQKNADQORFKTHV 815  
 QY 258 ETRFYVDYDDLMLSLKLG--AAKGMINTCEYQ 289  
 DB 816 Q-----NRKLSTLKGEDIVDMHLNTHQSQ 841  
 RESULT 40  
 HLYA\_ECOLI  
 ID HLYA\_ECOLI STANDARD; PRT; 1024 AA.  
 AC P08715;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Hemolysin, plasmid.  
 GN Name=hlyA;  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hesse J., Wells W., Vogel M., Goebel W.;  
 RT "Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparison with a corresponding chromosomal hemolysin sequence.";  
 RL PEMS Microbiol. Lett. 34:1-11(1986).  
 RN [2]  
 RP PALMITOYLATION OF LYS-564 AND LYS-690.  
 RX MEDLINE=9509325; PubMed=7801126;  
 RA Stanley P., Packman L.C., Koronakis V., Hughes C.;  
 RT "Fatty acylation of two internal lysine residues required for the toxic activity of Escherichia coli hemolysin.";  
 RL Science 266:1992-1996(1994).  
 RN [3]  
 RP PALMITOYLATION OF LYS-564 AND LYS-690.  
 RX MEDLINE=96404790; PubMed=8808931;  
 RA Ludwig A., Garcia F., Bauer S., Jarchau T., Benz R., Hoppe J., Goebel W.;  
 RT "Analysis of the in vivo activation of hemolysin (HlyA) from Escherichia coli.";

RL J. Bacteriol. 178:5422-5430(1996).  
 CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood  
 CC cell membranes and cause cell rupture by mechanisms not clearly  
 CC defined.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- DOMAIN: The Gly-rich region is probably involved in binding  
 CC calcium, which is required for target cell-binding or cytolytic  
 CC activity.  
 CC -!- DOMAIN: The three transmembrane domains are believed to be  
 CC involved in pore formation by the cytotoxin.  
 CC -!- PTM: Palmitoylated by hlyC. The toxin only becomes active when  
 CC modified.  
 CC -!- MISCELLANEOUS: The hemolysin of *E. coli* is produced predominantly  
 CC by strains causing extraintestinal infections, such as those of  
 CC the urinary tract.  
 CC -!- SIMILARITY: Belongs to the RTX prokaryotic toxin family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M14107; AAA98233.1; --  
 CC InterPro; IPR001343; Hemlysn\_Ca\_bind.  
 CC InterPro; IPR003995; RtxA.  
 CC InterPro; IPR011049; Serralyzn\_like\_C.  
 CC Pfam; PF00353; HemolysinCbind; 6.  
 CC Pfam; PF02382; RTX; 1.  
 CC PRINTS; PR00313; CABNDNGRPT.  
 CC PRINTS; PR01488; RTXTOXINA.  
 CC PROSITE; PS00330; HEMOLYSIN\_CALCIIUM; 4.  
 CC Calcium; Cytolysin; Hemolysis; Lipoprotein; Palmitate; Plasmid;  
 CC Repeat; Toxin; Transmembrane.  
 CC TRANSMEM 238 260 Potential.  
 CC TRANSMEM 268 327 Potential.  
 CC TRANSMEM 365 411 Potential.  
 CC DOMAIN 724 870 16 X REPEATS, GLY-RICH.  
 CC REPEAT 724 729 1.  
 CC REPEAT 733 738 2.  
 CC REPEAT 742 747 3.  
 CC REPEAT 751 756 4.  
 CC REPEAT 760 765 5.  
 CC REPEAT 769 774 6.  
 CC REPEAT 778 783 7.  
 CC REPEAT 787 792 8.  
 CC REPEAT 796 801 9.  
 CC REPEAT 807 812 10.  
 CC REPEAT 817 822 11.  
 CC REPEAT 826 831 12.  
 CC REPEAT 835 840 13.  
 CC REPEAT 844 849 14.  
 CC REPEAT 856 861 15.  
 CC REPEAT 865 870 16.  
 CC LIPID 564 564 N(6)-palmitoyl lysine.  
 CC LIPID 690 690 N(6)-palmitoyl lysine.  
 CC SEQUENCE 1024 AA; 110201 MW; 83944917F76C945B CRC64;  
 CC  
 CC Query Match 7.5%; Score 112.5; DB 1; Length 1024;  
 CC Best Local Similarity 20.5%; Pred. No. 71;  
 CC Matches 69; Conservative 55; Mismatches 116; Indels 97; Gaps 11;  
 CC  
 CC 2 TGFAETVEVK-----SAIETADGALDFYKVLQVDPKTFDEIKELSRKOE 53  
 CC 155 TALSSMKIDELIKKQSGGNGVSSSELAKEIQLNQLVDTVASLN-----NNVNSFSQ 208  
 CC 54 YSQEASVY-----VGDIKVLIMDSQDKYFEATQTVVEWCGVWVOLLISAYILL-FDEY 104  
 CC 209 LNTLGSVLSNTHKLVNGV-KLQNLPLNDIGAGLDTV---SGILSAIGASFILSNAD 264  
 CC 105 NEKASAKQDILIRLDD---GVNKLNEAKSLILGSSQSFNNASGKLLALDLSQLTDFSE 161

Db 265 TRTKAAAGVELTTKVLGNVGKGISQYIIAQRAAQLST----- 303  
 Qy 162 KSSYFOSQVDRIRKEAYAGAAAGIVAGPFLIIS-----YSIAAGVIEGKLIPELNDRLKA 217  
 Db 304 -----AAAGLIASAVTLAISPLSFLSIADKFRANKIEYSORFKK 345  
 Qy 218 VQNFFTSLSVTVKQANKIDAAKLIKLAETIAIGETETRTTFRFYVDYDDMLSLKGA 277  
 Db 346 LGVDGSLLAAPHKETGAIDASLTISTVLASVSSGISAATSLV---GAPVSALVGA 401  
 Qy 278 -----AKMINTCNEYQORHCK 294  
 Db 402 VTGIISGLEASKQAMFEHVASKQADVIABWEKKHCK 438  
 RESULT 41  
 OS0733 PRELIMINARY; PRT; 1098 AA.  
 ID OS0733;  
 AC OS0733;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Hypothetical protein BBG10.  
 GN OrderedLocustNames=BBG10;  
 OS Borrelia burgdorferi (lyme disease spirochete).  
 OG Plasmid lp28-2.  
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.  
 OC NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;  
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
 RA Lachiga R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,  
 RA Gwin M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,  
 RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,  
 RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,  
 RA Gocayne J.D., Weidman J.F., Utterback T.R., Matthey L., McDonald L.A.,  
 RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,  
 RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;  
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
 RT burgdorferi";  
 RL Nature 390:580-586(1997).  
 DR EMBL; AB000786; AAC66075.1; --  
 DR PIR; B70232; B70232.  
 DR TIGR; BBG10; --  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0004871; F:signal transducer activity; IEA.  
 DR GO; GO:0006935; P:chemotaxis; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR InterPro; IPR004089; Chmtaxis\_transd.  
 DR Pfam; PF00015; MCPsignal; 1.  
 KW Complete proteome; Hypothetical protein; Plasmid.  
 SQ SEQUENCE 1098 AA; 123861 MW; F465D18421F05935 CRC64;  
 Query Match 7.5%; Score 112.5; DB 2; Length 1098;  
 Best Local Similarity 20.6%; Pred. No. 77;  
 Matches 79; Conservative 60; Mismatches 122; Indels 123; Gaps 17;  
 Qy 10 VEVVKSALJETADGALDFYKVK-----LDQVPMKTFDEIKELSR 49  
 Db 552 VQVEKSVNFEFNSVDFVNEYNQLLKEKESREREIKTLPHTDQVSALQKLNDINEKNK 611  
 Qy 50 -FKQYSQEASVLYGDIK--VLLMDSQDKYFEATQ-----TVVEWCGV 90  
 Db 612 AFVEYKGSFETLNESSNQVVALEKQVNEKTDALDRSFVEAKALQKEITDLEW---E 668  
 Qy 91 TOLLISAYILLFDEYNEKASAKQDILIRI-----LDD---GVNKLNEAK- 132  
 Db 669 TWLLPA-----KERASAEKQWASKIQAMYKFFVDEHKSQPKLNETNRTIKQYAE 719  
 Qy 133 -----SLGSSQSFNNA-----SGKLLALDS-----QLTNDSEKSSYFSQV 170

```

Db 720 KAQDTTYSYDSMIDGLANFYKNAFMKDIAGKFLNKDTGESIGEEFNLNGKDVNNGEGL 779
QY 171 DRIRKEAYAG-----AAAGIVAGPGLIISYSIAA-----GVIRGKLIPDLNDRKAV 218
Db 780 EXWTTOMYESWKTGLTAGAGVFGPGWEAVALINGLTDFFVWGLKG-----QEKARIKAI 835
QY 219 QNFFTSLSVTVKQANKDIDAAKLKATEIAAI-----GEIKTETETFRFYVDYDDMLSL 273
Db 836 E-----KKRDEDLSEKRSVVELKLEDRPDEBIKMKRKEKLSLDDEYTKIEIF 885
QY 274 LKNAKKMINTCNEYQOR-HGKKT 296
Db 886 LKQAGSQGQISGEEFQKRLHVDQT 909

RESULT 42
ID Q6C359 PRELIMINARY; PRT; 1906 AA.
AC Q6C359;
DT 01-OCT-2004 (TREMBLrel. 28, Created)
DT 01-OCT-2004 (TREMBLrel. 28, Last sequence update)
DE Similar to DEHAOC09658g Debaryomyces hansenii IPF 1836.1.
GN ORFNames=YALIOF02387g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG GENOLEVURES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangoul L., Aigle M., Anthonard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckrich J.M., Beyne E., Bleykasten C.,
RA Boissame A., Boyer J., Cattolico L., Confantolero F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Kozul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissbach J.,
RA Winkler P., Souciet J.L.;
FT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382132; CAG77705.1; -.
SQ SEQUENCE 1906 AA; 210346 MW; F6ED7D1AF7B7562B CRC64;

Query Match
Best Local Similarity 22.1%; Pred. No. 1.5e+02;
Matches 66; Conservative 49; Mismatches 115; Indels 69; Gaps 12;

QY 7 EOTVEVKSATAGALDFYKNYL-----DQVIPWKTPTDETIKELSRFKQYSQBSVLIV 62
Db 1195 EKELETKTSELETKTAELTKKELTAKSDEA---TTYSAKVKELETSAALEKQTTLK 1251
QY 63 GDIKVLMDSQKPYEATQVYEWCGVVTQLLSAYILLFDEYNKKSQAQKIDILRLDD 122
Db 1252 AMADNLTQDLAEK-----TKELVAAKSEL---ESSNTSSKEEDVLTQKLSLD 1295
QY 123 GYNKLENAOKSLGSSQSFNNASGKLALDLSOLTNDPSEKSSYFOSQVDRIKEAVAGAA 182
Db 1296 ATAEEVELKKS---SQAAETEASSKVSLEAKLTK-ASESS---KAELEDKVNK----- 1341

```

```

QY 183 AGIVAGPGLIISYSIAAGVIE-----GKLIPELNDRLKAVQNF---FTLSVTVTKQA 232
Db 1342 -----LSSFKELQTSKDDHSTSVKSLTEQVRESTLKAENFEHDISSLKDLAQ 1392
QY 233 NKDIDAAKLKATEIAAIGEIKTETETFRFYVDYDDMLSLKGAKKMINTCNEYQOR 291
Db 1393 EKERDALRTELDTSIK-----EMENERTSLTKDAD-----SATKELTNKVSMLQTK 1438

RESULT 43
O45614 PRELIMINARY; PRT; 3102 AA.
AC O45614; P91824; Q9TZR4;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE C. elegans LAM-3 protein (Corresponding sequence T22A3.8) (Laminin
DE alpha chain) (Hypothetical protein T22A3.8).
GN Name=lam-3; Synonyms=lamal/2; ORFNames=T22A3.8;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Barlow K.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=22668894; PubMed=12783803;
RA Huang C.C., Hall D.H., Hedgecock E.M., Kao G., Karantz V.,
RA Vogel B.E., Hutter H., Chisholm A.D., Yurchenco P.D., Wadsworth W.G.;
RT "Laminin alpha subunits and their role in C. elegans development.";
RL Development 130:3343-3358 (2003).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurray A.A.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL008585; CAA15432.3; -.
DR EMBL; Z81125; CAA15432.3; JOINED.
DR EMBL; AL008585; CAB03385.3; JOINED.
DR EMBL; Z81125; CAB03385.3; -.
DR EMBL; AF074902; AAC26793.1; -.
DR FIR; F87908; F87908.
DR FIR; T23064; T23064.
DR FIR; T43291; T43291.
DR HSSP; P02468; INPE.
DR WormPep; T22A3.8; CE31067.
DR GO; GO:0005578; C:extracellular matrix; IEA.
DR GO; GO:0005606; C:laminin-1; IEA.
DR GO; GO:0005103; F:receptor binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030155; P:regulation of cell adhesion; IEA.
DR GO; GO:0030334; P:regulation of cell migration; IEA.
DR GO; GO:0045995; P:regulation of embryonic development; IEA.
DR InterPro; IPR008985; Cona_like_lec_g1.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR009030; Grow fac recept.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.

```

InterPro; IPR009254; Laminin I.  
 InterPro; IPR010307; Laminin II.  
 InterPro; IPR008211; Laminin N.  
 InterPro; IPR003129; TSP N.  
 Pfam; PF00052; Laminin B; 2.  
 Pfam; PF00053; Laminin\_EGF; 17.  
 Pfam; PF02210; Laminin\_C2; 3.  
 Pfam; PF06008; Laminin I; 1.  
 Pfam; PF06009; Laminin II; 1.  
 Pfam; PF00055; Laminin N; 1.  
 PRINTS; PR00011; EGF\_LAMININ.  
 ProDom; PD003031; Laminin B; 2.  
 SMART; SM00180; EGF\_Lam; 17.  
 SMART; SM00281; LamB; 2.  
 SMART; SM00282; LamG; 4.  
 SMART; SM00136; LamNT; 1.  
 PROSITE; PS00022; EGF 1; 12.  
 PROSITE; PS01186; EGF 2; 1.  
 PROSITE; PS01248; LAMININ\_TYPE\_EGF; 14.  
 PROSITE; PS00025; LAM G\_DOMAIN; 3.  
 Hypothetical protein; Laminin EGF-like domain.  
 KW Hypothetical protein; Laminin EGF-like domain.  
 SQ SEQUENCE 3102 AA; 343790 MW; 2EE9305DC591FAB CRC64;

Query Match 7.5%; Score 112.5; DB 2; Length 3102;  
 Best Local Similarity 19.9%; Pred. No. 2.6e+02;  
 Matches 73; Conservative 64; Mismatches 106; Indels 123; Gaps 16;

QY 6 ABOQTEVW-----KSAIETADGALDFYNNKLDQVDPKWTDETTIKELSRFKQEQYSQAS 59  
 Db 1862 ABEELQWVTAQKLNITFDLKNRIDVLEQWMD-----YRETIYDVSKDQADARMS 1915

QY 60 VLVG-----DIKVLMDSDQK-----YFEATQTV----- 83  
 Db 1916 LVVGKINRYKEVSNIEKRLVAREDOAIYRSNSIEKARSELMMWFDEKKEKINTLAEL 1975

QY 84 ---YEWCGVVTQLLSAYILLFFDEYNEK-----KASAQKDILIRILDGQVKNLN 128  
 Db 1976 PDLVEQCNITLL---YSQLIDYDEYVYVOTAGRAHAEKLEVOAQK-IVDRFVDVTRTETEN 2031

QY 129 -----BAQKS-----LLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRI 173  
 Db 2032 PLKASHAYENIVEALKNATEVDSAAEASVSKMLGSGSGSGDANEES--LRSLQELK 2089

QY 174 RKAAYAGAAGIVAGPFGLLIYSIAAGVIE--GKLIPELNDRLKAVQNFFTSLSVTVKQ 231  
 Db 2090 KNESLSN-----VNSNAVKIVEELKKEKDLTRLGHNLKTSI----- 2131

QY 232 ANKDDIAAKLKLATEIAAIGEIKTETTRFYVDYDDMLSLKGAQKQ-INTCNEYQQ 290  
 Db 2132 -----VKRLGVKNEASS---WDDKDRMHSLKNGAKTAHERSANVCKE 2173

QY 291 RHGKKT 296  
 Db 2174 SEGIKT 2179

RESULT 44  
 Q6QNE7 PRELIMINARY; PRT; 1012 AA.  
 ID Q6QNE7  
 AC Q6QNE7;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OC NCBI\_TaxID=8355;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE=Spleen;  
 RX MEDLINE=22388257; PubMed=12477932;

Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner B., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J., Marra M.A.;  
 "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [2]  
 SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;  
 RX MEDLINE=22341132; PubMed=12454917;  
 RA Klein S.L., Strausberg R.L., Wagner R.L., Pontius J., Clifton S.W., Richardson P.;  
 RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative.";  
 RL Dev. Dyn. 225:384-391(2002).  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE=Spleen;  
 RA Klein S., Strausberg R.;  
 RA Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC073565; AAH73565.1; -  
 DR InterPro; IPR009053; Prefoldin.  
 DR InterPro; IPR007794; Rib\_recept\_KP.  
 DR Pfam; PF05104; Rib\_recP\_KP\_reg; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1012 AA; 112965 MW; CF23AB8E0D9EDAEEA CRC64;

Query Match 7.4%; Score 112; DB 2; Length 1012;  
 Best Local Similarity 20.7%; Pred. No. 75;  
 Matches 68; Conservative 61; Mismatches 127; Indels 72; Gaps 14;

QY 8 QTVEVKSATFADGALDFYNNKLD-QVTP-----KTFETIKELSRFKQEQYSQ 56  
 Db 665 QIASIQKEAELKEAVEQCKNNNDLREKNQWQAMEALGLAEKCEKLNSEKAKEEMVQ 724

QY 57 EASVLVGDIKVLMD-----SODKYFEATQTVYEWCGVVTQLLSAYILLFFDEYNEKK 108  
 Db 725 QLSAVQSQTKELOSPLPQITIVSQSYSEWLQ---EPRFTSQLS-----QOTEKE 774

QY 109 ASAQKDILIRILDGQVKNLN---BAQKSLLGSSQSFNNASGKLLALDSQLTNDPSEKSSY 165  
 Db 775 GSSELOLQKQEDSQSDLOVECEKYRTILQTEAM-----LKALQNSVEEEQVWKAK 828

QY 166 FQSQVDRIKRAYAGAAAGIVAGPFGLLIYSIAAGVIEGKLIPELNDRLKAVQNFFTSL 225  
 Db 829 FSSSEELR-----SHSQVKTLEE---TVKLRDLDSTQOLKECV 867

QY 226 SVTVKQANKDIDAAKILKLT---ETAAIGEIKTET-----ETTRFYVDYDDMLSLKGA 278  
 Db 868 SLMEAQLETQWNAKTECQTVSNELISQQLSSEHRLDATKAEARQSIELSIVLRQOL 927

QY 279 KKMINTCNEYQQRHG---KKTL--LEVP 301  
 Db 928 GEMLNHVND-TEKHGTETEHQTLQNEKVP 954

RESULT 45  
 Q7RFQ7 PRELIMINARY; PRT; 1081 AA.  
 ID Q7RFQ7





Cell 42:173-182(1985).  
-!- FUNCTION: Not known but may be related to activation of the  
variant surface glycoprotein genes.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M11452; AAA30191.1; -  
PIR; A03395; VMUT21.  
DR InterPro; IPR004922; ESAG.  
DR Pfam; PF03238; ESAG1; 1.  
KW Signal.  
FT SIGNAL 1 23  
FT CHAIN 24 329  
VSG expression site-associated protein  
221A.  
FT CARBOHYD 73 73 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 294 294 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 308 308 N-linked (GlcNAc...) (Potential).  
FT SEQUENCE 329 AA; 36603 MW; 4D19F59477D9CEE8 CRC64;  
SQ  
Query Match 7.4%; Score 111; DB 1; Length 329;  
Best Local Similarity 22.9%; Pred. No. 23;  
Matches 48; Conservative 33; Mismatches 63; Indels 66; Gaps 9;  
QY 92 QLLSAVILLFD-----EYNEKKAQAQKDLIRILDGYNKL-NEAOKS 133  
DB 4 EIVELVLLFSVTCVDAMLOGADCTRVADHKEHAPVTEAVCYLRCLSDALNKLYSEGEKK 63  
QY 134 LIGSSQSFNNAGKLLALDSQLTNDPSEKSY-----FQSDVRIRKEAYAGAAAGIV 186  
DB 64 LLVTEEVYANAS---LILD-DMEGRAGESSTYLSVIRGVMEEQTDRLK- 108  
QY 187 AGPFGLLISYSIAAGVI---EGKLIPELNDRLKAVQ-----NFTSLSVTVKQA 232  
DB 109 -----LISYGNKGNLVAKAGLFAALDSLSKEVRKEIPGALIKNTKNTYTSVAEIVRTV 162  
QY 233 NKDIDAARKL-----ATEIAAIGEIKTE 256  
DB 163 WEDVGEILWKETEAKCGSKQVGEVGEIQTE 192  
RESULT 48  
Q6ZM87 PRELIMINARY; PRT; 329 AA.  
AC Q6ZM87  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE ESAG1.  
GN Names=H25N7.33; Synonyms=ESAG1;  
OS Trypanosoma brucei.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5691;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=427;  
RX MEDLINE=98434256; PubMed=9763292;  
RA Rudenko G., Chaves I., Dirks-Mulder A., Borst P.;  
RT "Selection for activation of a new variant surface glycoprotein gene  
expression site in Trypanosoma brucei can result in deletion of the  
old one.";  
RT old one.";  
RL EMBL; AL671259; CAD21890.1; -  
DR EMBL; AL671259; CAD21890.1; -  
SQ SEQUENCE 329 AA; 36603 MW; 4D19F59477D9CEE8 CRC64;  
[2]  
SEQUENCE FROM N.A.  
RC STRAIN=427;  
RX MEDLINE=22103446; PubMed=12106867;  
RA Berriman M., Hall N., Shearer K., Bringaud F., Tiwari B., Isobe T.,  
Bowman S., Corton C., Cross G.A.M., Hoek M., Zanders T.,

Barberof M., Borst P., Rudenko G.;  
RT "The architecture of variant surface glycoprotein gene expression  
sites in Trypanosoma brucei.";  
RL Mol. Biochem. Parasitol. 122:131-140(2002).  
[3]  
RN SEQUENCE FROM N.A.  
RP STRAIN=427;  
RC STRAIN=427;  
RA Zeng C., Zhao B., Hiehl M., Catanese J., Gerard C., Melville S.E.,  
Hoek M., Navarro M., Cross G.A.M., El-Sayed N., Barberof M.,  
Rudenko G., Borst P., de Jong P.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL671259; CAD21890.1; -  
DR InterPro; IPR004922; ESAG.  
DR Pfam; PF03238; ESAG1; 1.  
DR SEQUENCE 329 AA; 36603 MW; 4D19F59477D9CEE8 CRC64;  
SQ  
Query Match 7.4%; Score 111; DB 2; Length 329;  
Best Local Similarity 22.9%; Pred. No. 23;  
Matches 48; Conservative 33; Mismatches 63; Indels 66; Gaps 9;  
QY 92 QLLSAVILLFD-----EYNEKKAQAQKDLIRILDGYNKL-NEAOKS 133  
DB 4 EIVELVLLFSVTCVDAMLOGADCTRVADHKEHAPVTEAVCYLRCLSDALNKLYSEGEKK 63  
QY 134 LIGSSQSFNNAGKLLALDSQLTNDPSEKSY-----FQSDVRIRKEAYAGAAAGIV 186  
DB 64 LLVTEEVYANAS---LILD-DMEGRAGESSTYLSVIRGVMEEQTDRLK- 108  
QY 187 AGPFGLLISYSIAAGVI---EGKLIPELNDRLKAVQ-----NFTSLSVTVKQA 232  
DB 109 -----LISYGNKGNLVAKAGLFAALDSLSKEVRKEIPGALIKNTKNTYTSVAEIVRTV 162  
QY 233 NKDIDAARKL-----ATEIAAIGEIKTE 256  
DB 163 WEDVGEILWKETEAKCGSKQVGEVGEIQTE 192  
RESULT 49  
CAD21890 PRELIMINARY; PRT; 329 AA.  
ID CAD21890  
AC CAD21890;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
DE ESAG1.  
GN H25N7.33.  
OS Trypanosoma brucei.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5691;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=427;  
RX MEDLINE=98434256; PubMed=9763292;  
RA Berriman M., Hall N., Shearer K., Bringaud F., Tiwari B., Isobe T.,  
Bowman S., Corton C., Cross G.A.M., Hoek M., Zanders T.,  
Barberof M., Borst P., Rudenko G.;  
RT "The architecture of variant surface glycoprotein gene expression  
sites in Trypanosoma brucei.";  
RL Mol. Biochem. Parasitol. 122:131-140(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=427;  
RX MEDLINE=98434256; PubMed=9763292;  
RA Rudenko G., Chaves I., Dirks-Mulder A., Borst P.;  
RT "Selection for activation of a new variant surface glycoprotein gene  
expression site in trypanosoma brucei can result in deletion of the  
old one.";  
RT old one.";  
RL EMBL; AL671259; CAD21890.1; -  
DR EMBL; AL671259; CAD21890.1; -  
SQ SEQUENCE 329 AA; 36603 MW; 4D19F59477D9CEE8 CRC64;  
[2]  
SEQUENCE FROM N.A.  
RC STRAIN=427;  
RX MEDLINE=22103446; PubMed=12106867;  
RA Berriman M., Hall N., Shearer K., Bringaud F., Tiwari B., Isobe T.,  
Bowman S., Corton C., Cross G.A.M., Hoek M., Zanders T.,



```
QY 92 QLLSAVILLFD-----EYNEKSAQKDIILIRLDDGVNKL-NEAOKS 133
DB 4 EIVELVLLFSVTCVDAWLQGDCTRVADHKEHAPVTEAVCYLRCLSDALNKLSEGEKK 63
QY 134 LGSSQSFNNASGKLLALDSQTNDFSEKSSY-----FQSOVDRIKREAVAGAAAGIV 186
DB 64 LLVTEEVANAS---LILD-DMEGRAGESSTVLSVIRGMEQTDRLK-----108
QY 187 AGPGLIISYSIAAGVI---EGKLIPELNDRILKAVQ-----NFFTSLSVTVKQA 232
DB 109 -----LISYGNKGMNLVAKAGGLFALEDSLKEVRKEIPGALIKNTKYTTSVAEIVRTV 162
QY 233 NKDIDAAKLKL-----ATEIAAIGEIKTE 256
DB 163 WEDVGEILWKETEAKCGSQKVEGVGEIQTE 192

RESULT 50
Q73AJ3 PRELIMINARY; PRT; 465 AA.
AC Q73AJ3;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Flagellin.
GN OrderedLocusNames=BCE1780;
OS Bacillus cereus (strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=222523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714;
RA Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,
RA Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RT adaptations and a large plasmid related to Bacillus anthracis pXOI.";
RL Nucleic Acids Res. 32:977-988(2004).
DR EMBL; AE017269; AAS40708.1; -.
DR TIGR; BCE1780; -.
DR InterPro; IPR001029; Flagellin_C.
DR InterPro; IPR001492; Flagellin_N.
DR Pfam; PF00700; Flagellin_C; 1.
DR Pfam; PF00669; Flagellin_N; 1.
DR PRINTS; PR00207; FLAGELLIN.
DR ProDom; PD000316; Flagellin_C; 1.
KW Complete proteome; Flagellum.
SQ SEQUENCE 465 AA; 49855 MW; E4C307E7717381D5 CRC64;

Query Match 7.4%; Score 111; DB 2; Length 465;
Best Local Similarity 22.0%; Pred. No. 35;
Matches 70; Conservative 51; Mismatches 129; Indels 68; Gaps 14;

QY 6 ABQTEVVVKSAIETADGALDFYNKYLDQV---IPWKTDETIKELSRPK-----Q 52
DB 184 AKATI-TAKDAPDAKADAAKEIDAAAKDTPSKNDQAQIAKEYIEAKATLATLKPTDA 242
QY 53 EYSQEAASVL-----VGDIKVLMDSDQKYPEATQVYEWCGVVTQLLSAVILLFDEYNE 106
DB 243 TYAKAAELDRAATNALNDNAKVLVDGYEKKLTITTKKEA-----EYTA 285
QY 107 KKASAKQDILIRLDDGVNKLNEAQKSLGSSQSFNNASGKLLALDSQTNDFSEKSSYP 166
DB 286 AKEQSTKSTAAA---DLVTKYETAKSNALGNDIAKEYLEAK-TAVEAN-KNDISSKSR-F 339
QY 167 QSOVDRIKREAVAGAAAGIVAGPGLIISYSIA--AGVIEGKL--IPELNDRILKAVQNPFF 222
DB 340 EAAETELNKDITANKAAKY-----LVETYEKAKTAGTTEKSLVAVDKIDEALKTIADNR 393
QY 223 TSLSVTVKQANKDIDAAKCLKATEIAAIGEI-----KTETETTRFYVDYDDMLSLKLG 276
DB 394 ATLGATLNRLDPNVNNLKSQSASMASAQIEDADMAKEMSEMTKF-----KILNE 444
```

QY 277 AAKKMINTCNEYQQRHGK 294  
DB 445 AGISMLSQANQTPOQWVK 462

Search completed: January 28, 2005, 19:16:04  
Job time : 205 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:16:10 ; Search time 143 Seconds  
(without alignments)

765.529 Million cell updates/sec

Title: US-09-993-292B-24

Perfect score: 1510

Sequence: 1 MTGIPAEQTEVVKSAIETA.....TCNEYQQRHGKKTLLVDPDI 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 1466  | 97.1        | 305    | 9  | US-09-993-292A-2     |
| 2          | 136.5 | 9.0         | 889    | 16 | US-10-437-963-197045 |
| 3          | 116   | 7.7         | 891    | 14 | US-10-369-493-10918  |
| 4          | 116   | 7.7         | 941    | 15 | US-10-282-122A-46343 |
| 5          | 116   | 7.7         | 1023   | 10 | US-09-884-696-5      |
| 6          | 115   | 7.6         | 1005   | 14 | US-10-369-493-1061   |
| 7          | 112.5 | 7.5         | 2823   | 14 | US-10-369-493-5220   |
| 8          | 112.5 | 7.5         | 2823   | 14 | US-10-369-493-5221   |
| 9          | 111   | 7.4         | 1295   | 14 | US-10-369-493-6440   |
| 10         | 110.5 | 7.3         | 564    | 15 | US-10-424-599-212070 |
| 11         | 110   | 7.3         | 1189   | 15 | US-10-282-122A-70920 |
| 12         | 108.5 | 7.2         | 1959   | 14 | US-10-028-248A-106   |
| 13         | 108.5 | 7.2         | 1959   | 15 | US-10-107-782-106    |

|    |       |     |      |    |                      |                    |
|----|-------|-----|------|----|----------------------|--------------------|
| 14 | 108.5 | 7.2 | 1999 | 14 | US-10-028-248A-107   | Sequence 107, App  |
| 15 | 108.5 | 7.2 | 1999 | 15 | US-10-107-782-107    | Sequence 107, App  |
| 16 | 108   | 7.2 | 815  | 16 | US-10-437-963-191043 | Sequence 191043, A |
| 17 | 108   | 7.2 | 1965 | 14 | US-10-369-493-3279   | Sequence 3279, Ap  |
| 18 | 107.5 | 7.1 | 742  | 15 | US-10-282-122A-53523 | Sequence 53523, A  |
| 19 | 107.5 | 7.1 | 857  | 14 | US-10-369-493-9374   | Sequence 9374, Ap  |
| 20 | 106.5 | 7.1 | 1938 | 14 | US-10-369-493-5109   | Sequence 5109, Ap  |
| 21 | 106.5 | 7.1 | 1938 | 14 | US-10-369-493-5110   | Sequence 5110, Ap  |
| 22 | 106.5 | 7.1 | 1962 | 15 | US-10-282-122A-71235 | Sequence 71235, A  |
| 23 | 106   | 7.0 | 1961 | 14 | US-10-028-248A-105   | Sequence 105, App  |
| 24 | 106   | 7.0 | 1961 | 15 | US-10-107-782-105    | Sequence 105, App  |
| 25 | 105.5 | 7.0 | 344  | 14 | US-10-369-493-497    | Sequence 497, App  |
| 26 | 105   | 7.0 | 737  | 16 | US-10-437-963-107270 | Sequence 107270, A |
| 27 | 104   | 6.9 | 722  | 9  | US-09-815-242-10796  | Sequence 10796, A  |
| 28 | 103.5 | 6.9 | 660  | 10 | US-09-841-260-139    | Sequence 139, App  |
| 29 | 103.5 | 6.9 | 660  | 13 | US-10-007-693-139    | Sequence 139, App  |
| 30 | 103.5 | 6.9 | 660  | 16 | US-10-762-058-139    | Sequence 139, App  |
| 31 | 103.5 | 6.9 | 1020 | 15 | US-10-282-122A-58016 | Sequence 58016, A  |
| 32 | 103.5 | 6.9 | 1163 | 15 | US-10-282-122A-51864 | Sequence 51864, A  |
| 33 | 103.5 | 6.9 | 5373 | 16 | US-10-408-765A-741   | Sequence 741, App  |
| 34 | 103   | 6.8 | 862  | 14 | US-10-369-493-19762  | Sequence 19762, A  |
| 35 | 103   | 6.8 | 961  | 14 | US-10-080-608A-66    | Sequence 66, Appl  |
| 36 | 103   | 6.8 | 961  | 14 | US-10-370-685-155    | Sequence 155, App  |
| 37 | 103   | 6.8 | 1583 | 16 | US-10-408-765A-1635  | Sequence 1635, Ap  |
| 38 | 102.5 | 6.8 | 472  | 15 | US-10-282-122A-55283 | Sequence 5283, A   |
| 39 | 102   | 6.8 | 590  | 15 | US-10-220-480-50     | Sequence 50, Appl  |
| 40 | 102   | 6.8 | 590  | 16 | US-10-220-481-155    | Sequence 155, App  |
| 41 | 102   | 6.8 | 612  | 15 | US-10-220-480-56     | Sequence 56, Appl  |
| 42 | 102   | 6.8 | 612  | 16 | US-10-220-481-161    | Sequence 161, App  |
| 43 | 102   | 6.8 | 1903 | 17 | US-10-766-993-3      | Sequence 3, Appli  |
| 44 | 102   | 6.8 | 1948 | 14 | US-10-032-585-7611   | Sequence 7611, App |
| 45 | 101.5 | 6.7 | 1938 | 14 | US-10-171-311-164    | Sequence 164, App  |
| 46 | 101.5 | 6.7 | 1945 | 10 | US-09-927-597-2      | Sequence 2, Appli  |
| 47 | 101.5 | 6.7 | 1972 | 14 | US-10-171-311-162    | Sequence 162, App  |
| 48 | 101.5 | 6.7 | 1972 | 14 | US-10-341-434-103    | Sequence 103, App  |
| 49 | 101.5 | 6.7 | 1979 | 10 | US-09-927-597-4      | Sequence 4, Appli  |
| 50 | 101   | 6.7 | 1475 | 14 | US-10-369-493-6362   | Sequence 6362, Ap  |
| 51 | 100.5 | 6.7 | 425  | 14 | US-10-434-418-3      | Sequence 3, Appli  |
| 52 | 100.5 | 6.7 | 526  | 14 | US-10-308-936-15     | Sequence 15, Appl  |
| 53 | 100.5 | 6.7 | 529  | 15 | US-10-282-122A-48338 | Sequence 48338, A  |
| 54 | 100.5 | 6.7 | 553  | 14 | US-10-308-936-14     | Sequence 14, Appl  |
| 55 | 100.5 | 6.7 | 609  | 14 | US-10-308-936-12     | Sequence 12, Appl  |
| 56 | 100.5 | 6.7 | 609  | 16 | US-10-408-765A-566   | Sequence 566, App  |
| 57 | 100.5 | 6.7 | 609  | 16 | US-10-408-765A-2133  | Sequence 2133, Ap  |
| 58 | 100.5 | 6.7 | 613  | 14 | US-10-013-477-18     | Sequence 18, Appl  |
| 59 | 100.5 | 6.7 | 613  | 14 | US-10-013-477-26     | Sequence 26, Appl  |
| 60 | 100.5 | 6.7 | 613  | 14 | US-10-207-655-204    | Sequence 204, App  |
| 61 | 100.5 | 6.7 | 613  | 14 | US-10-308-936-9      | Sequence 9, Appli  |
| 62 | 100.5 | 6.7 | 613  | 15 | US-10-284-237-2002   | Sequence 2002, Ap  |
| 63 | 100.5 | 6.7 | 613  | 16 | US-10-408-765A-571   | Sequence 571, App  |
| 64 | 100.5 | 6.7 | 613  | 17 | US-10-770-668-10     | Sequence 10, Appl  |
| 65 | 100.5 | 6.7 | 623  | 14 | US-10-308-936-11     | Sequence 11, Appl  |
| 66 | 100.5 | 6.7 | 627  | 14 | US-10-308-936-8      | Sequence 8, Appli  |
| 67 | 100.5 | 6.7 | 803  | 9  | US-09-815-242-5597   | Sequence 5597, Ap  |
| 68 | 100.5 | 6.7 | 805  | 9  | US-09-815-242-12286  | Sequence 12286, A  |
| 69 | 100.5 | 6.7 | 861  | 14 | US-10-310-154-485    | Sequence 485, App  |
| 70 | 100.5 | 6.7 | 861  | 14 | US-10-369-493-17514  | Sequence 17514, A  |
| 71 | 100.5 | 6.7 | 1009 | 15 | US-10-282-122A-43788 | Sequence 43788, A  |
| 72 | 100   | 6.6 | 451  | 14 | US-10-369-493-2102   | Sequence 2102, Ap  |
| 73 | 100   | 6.6 | 523  | 10 | US-09-976-782-6      | Sequence 6, Appli  |
| 74 | 100   | 6.6 | 680  | 15 | US-10-425-114-68086  | Sequence 68086, A  |
| 75 | 100   | 6.6 | 957  | 16 | US-10-437-963-176713 | Sequence 176713, A |
| 76 | 100   | 6.6 | 990  | 15 | US-10-108-260A-3896  | Sequence 3896, Ap  |
| 77 | 100   | 6.6 | 1174 | 14 | US-10-094-749-1911   | Sequence 1911, Ap  |
| 78 | 100   | 6.6 | 1939 | 17 | US-10-798-037-4      | Sequence 4, Appli  |
| 79 | 100   | 6.6 | 2481 | 16 | US-10-467-595-23     | Sequence 23, Appl  |
| 80 | 99.5  | 6.6 | 857  | 14 | US-10-369-493-9166   | Sequence 9166, Ap  |
| 81 | 99.5  | 6.6 | 1959 | 14 | US-10-028-248A-36    | Sequence 36, Appl  |
| 82 | 99.5  | 6.6 | 1959 | 15 | US-10-107-782-36     | Sequence 36, Appl  |
| 83 | 99.5  | 6.6 | 1961 | 14 | US-10-028-248A-103   | Sequence 103, App  |
| 84 | 99.5  | 6.6 | 1961 | 15 | US-10-107-782-103    | Sequence 103, App  |
| 85 | 99    | 6.6 | 268  | 9  | US-09-815-242-5337   | Sequence 5337, Ap  |
| 86 | 99    | 6.6 | 520  | 14 | US-10-369-493-1809   | Sequence 1809, Ap  |

Sequence 14486, A  
Sequence 44151, A  
Sequence 12211, A  
Sequence 17323, A  
Sequence 51602, A  
Sequence 4, Appli  
Sequence 248508,  
Sequence 286451,  
Sequence 11332, A  
Sequence 74458, A  
Sequence 115622,  
Sequence 63115, A  
Sequence 226623,  
Sequence 64494, A

ALIGNMENTS

RESULT 1  
US-09-993-292A-2  
; Sequence 2, Application US/09993292A  
; Patent No. US2002014630A1  
; GENERAL INFORMATION:  
; APPLICANT: James E. Gallen  
; APPLICANT: University of Maryland  
; TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF  
; TITLE OF INVENTION: PROTEINS  
; FILE REFERENCE: UOFMD.007A  
; CURRENT APPLICATION NUMBER: US/09/993,292A  
; CURRENT FILING DATE: 2001-11-23  
; PRIOR APPLICATION NUMBER: 60/252,516  
; PRIOR FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Salmonella Typhi  
US-09-993-292A-2  
Query Match 97.1%; Score 1466; DB 9; Length 305;  
Best Local Similarity 97.0%; Pred. No. 9e-118;  
Matches 294; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
QY 1 MTGIFABQTVVVKSAIETADGALDFYKYLQVVPWKTFTDETIKLSRFRQYSEASV 60  
DB 1 MTSIFAQTVVVKSAIETADGALDLYNKYLQVVPWKTFTDETIKLSRFRQYSEASV 60  
QY 61 LVGDIKVLMDSDQKYFEATQTVYEWCVVTVLLSAYILLFDEYNEKKSAAQKIDILIRIL 120  
DB 61 LVGDIKVLMDSDQKYFEATQTVYEWCVVTVLLSAYILLFDEYNEKKSAAQKIDILIRIL 120  
QY 121 DGVNKLNEAQSLLGSSQFNNAQKLLALDSQLTNDPSEKSSYFQSOVDRIKEAYAG 180  
DB 121 DGVNKLNEAQSLLGSSQFNNAQKLLALDSQLTNDPSEKSSYFQSOVDRIKEAYAG 180  
QY 181 AAGIVAGPGLIISYIAAGVIEGLIPELNDRLKAVQNFSTLSVTVKQANKDIDAAK 240  
DB 181 AAGIVAGPGLIISYIAAGVIEGLIPELNDRLKAVQNFSTLSVTVKQANKDIDAAK 240  
QY 241 LKLAETIAAIGETKTETETFRFVYDDMLSLKGAAKMINTCNEYQORHGKKTLLFV 300  
DB 241 LKLAETIAAIGETKTETETFRFVYDDMLSLKGAAKMINTCNEYQORHGKKTLLFV 300  
QY 301 PDI 303  
DB 301 PDV 303  
RESULT 2  
US-10-437-963-197045  
; Sequence 197045, Application US/10437963

Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbaruk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 197045  
; LENGTH: 889  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_92840C.1.pcp  
US-10-437-963-197045  
Query Match 9.0%; Score 136.5; DB 16; Length 889;  
Best Local Similarity 22.4%; Pred. No. 0.011;  
Matches 77; Conservative 67; Mismatches 121; Indels 79; Gaps 17;  
QY 8 QTVVVKSAIETADGALD-----FYNKYLDQVVPWKTFTDETIK-----ELSRFRQY 54  
DB 268 EKVEILSSEVVRLKGLDSTABSESKNRTETELV--KNLESEVSVLKGLEEARIEER 325  
QY 55 SQEASVLVGDIKVLMDSDQKYFEATQTVYEW---CGVTVQLLSAYILLFDEY-NEKKA 110  
DB 326 LAETEKLEIELKSEVADAKKAESEARQLFEWKHKAGLEMELEA-VTSDKPKGESLAS 384  
QY 111 AQDILIRILDDGVNKLNEAQSLLGSSQFNNAQKLLALD----SOLTNDPSEKSSYF- 166  
DB 385 TTEE-----LGKIQSALQDRESEIEVLKGTALTEIEVARLLADVNESNQFD 432  
QY 167 -----QSQVDRIKEAYAGAAAGIVAGPGLIISYIAAGV-IEG-----KLIPE 210  
DB 433 ASQEVFGLQTTIDVLRNKLKLEAAEASEA-----LNNEKAAVVKIEGLTEENVKLI 486  
QY 211 LND-----RLKAVQNFSTLSVTVKQANKDIDAAKILATE-----IAAIGETKTET 257  
DB 487 LNETDREKEKRAVEDLTAALS---EESDKAKEAHERYLSKEDDHEHALAQIGDKMAL 543  
QY 258 ETTR-----FYVDYDDMLSLKGAAKMINTCNEYQOR-HGKKT 296  
DB 544 KSTKESYEVMLDEANYDITCLRKKNVDKLEAEVKNKYRECESEKET 587  
RESULT 3  
US-10-369-493-10918  
; Sequence 10918, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 10918  
; LENGTH: 891

```

; TYPE: PRT
; ORGANISM: Ferroplasma acidarmanus
US-10-369-493-10918

Query Match 7.7% Score 116; DB 14; Length 891;
Best Local Similarity 22.3%; Pred. No. 0.62; Indels 66; Gaps 12;
Matches 69; Conservative 46; Mismatches 129;

QY 4 IFABQTEVVVKSATETADGALDFYNNKLDQOVI-----PWKTFDETIKELS 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 505 IFGSQIETAAASYNEIK-SLETEKKELENTIENASKGHELFNVLSAENENLEKAIKELQ 563
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 49 RFKQESQESASVLGDIKVLLMD--SODKFEAT-QTVYEWCGVVTQLLSAYILLFDEYN 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 564 QYENEVIRYSIIISG-----MDPGAIKKGIEAAGESFTTFKKNRINELLS-QIGFVPEYN 616
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 106 EKASQAQKQILIRILDDGVNKLNEAOKSLGSSQFNASGKLALDSQLNDFSEKSSY 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 EYQNTGKISSEINRLKTEVERSRBMKSLSIKDEIENRKSISGLRIEMENKQSAHQY 676
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 FQSQVDRIRKEAYAGAAAGIVAGPFGLIHSYSIAAGVIEGKLIPELNDRLK---AVQNFF 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 677 -----DGDIDQ-----AGNIESRYKSAYQENIKMKTLDVSYT 708
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 223 TSLSVTVKQA-NKDDIDAAKLKATE-IAATGEIKTETETTFRYVDYDDLMLSLKGAKK 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 ERIETEENAKNLEQDAEKYKKTREAITLTKI-----REAFDNGIQSIIRKDSAS 761
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 281 MINTCNEYQQ 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 762 MTNLTRKYLQ 771
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-10-282-122A-46343
; Sequence 46343, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308

```

```

; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46343
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-10-282-122A-46343

Query Match          7.7%; Score 116; DB 15; Length 941;
Best Local Similarity 20.1%; Pred. No. 0.67; 88; Indels 62; Gaps 7
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7

QY 7 ECTVEVVKSAIETADGAL-DFYNKYLDQVIP-WKTFDETIKELSRFKQESQEASVLVGD 64
Db      ::::: ||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
QY 596 KEWSEINEATONASAOQLGDLATYDSEIVPENTAIERTKMSKNTSQILKEADKKLPD 655
Db      ::::: ||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
QY 65 IKVLLMDSQDKYPEATQTVYEWGCVVTTOLLSAYILLFDEYNEKASAKOILIRILDGV 124
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 656 VKKLEBDSK-----GLV-----DGR 671
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 125 NKLNEAQKSLGSSQSFNNASGKLLALDSO-----LTNDFSEKSSYFQSOVDRIK 175
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 672 KKLADIKAEMPATEKKIKELADKIRDFESEEDUKDIIRLLXNDVEKSDYFANPNLKEN 731
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 176 EAYAGAAAGIVAGPFGFLIIISYIAAGVIEKGLIPELND---RLKAVQNFF----TSLSVT 228
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 732 KLFAMENYGSAMSPFYTVLALWVGALLWVSLLTVEVHEEGANYKSHEIYFGRLLTFLTMG 791
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 229 VKQA 232
Db      ::|||

QY 792 LSQA 795
Db      ::|||

RESULT 5
US-09-884-696-5
; Sequence 5, Application US/09884696
; Publication No. US20030035809A1
; GENERAL INFORMATION:
; APPLICANT: GEORGE, LISLE W
; APPLICANT: ANGELOS, JOHN A
; APPLICANT: HESS, JOHN F
; TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES
; TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA
; TITLE OF INVENTION: BOVIS INFECTIONS
; FILE REFERENCE: 481.06
; CURRENT APPLICATION NUMBER: US/09/884,696
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-884-696-5

Query Match          7.7%; Score 116; DB 10; Length 1023;
Best Local Similarity 20.2%; Pred. No. 0.75; 112; Indels 98; Gaps 11
Matches 68; Conservative 59; Mismatches 112; Indels 98; Gaps 11

QY 2 TGI FARQTVEVK-----SAIETADGALDFYNKYLDQVIPWKTFDETIKELSRFKQE 53
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 54 YSQEASVL-----VGDIKVLMDSQDKYFATQTVYEWGCVVTTOLLSAYILL-FDEY 104
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 208 LNKLGSLVNTKHLNGVGN-KLQNLPLNDIGAGLDTV---SGILSAISAFILSNADAD 263
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 105 NEKASAKOILIRIIDD---GVNKLNEAQKSLGSSQSFNNASGKLLALDSQLTNDPSE 161
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 264 TGTAAAGVELTTKVLGNVCKGISQVILIAORAAQGLSTS-----DGR 302
Db      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

162 KSSYFQSQVDRIKEAVAGAAAGVAGPFGLLIS-----YSIAAGVIRGKLIPELNDRLKA 217  
Db 303 -----NAAAGLIASVVTIAISPLSFLSIADKFRANKIEEYSORFKX 344  
Qy 218 VONFTSLSVTVKQAKDIDAUKLATEIAAIGEIKTETETTFYVDYDDLMLSLKGA 277  
Db 345 LGYDGSLLAAPHKETGAIDASLRISVTLASVSSGISAATTSLV-----GAPVSALVGA 400  
Qy 278 -----AKKMINTCNEYQQRHGK 294  
Db 401 VTGIIISGILEASKQAMFEHVASKWADVIAEWKXHGK 437  
RESULT 6  
US-10-369-493-1061  
; Sequence 1061, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 1061  
; TYPE: PRT  
; ORGANISM: Methanococcus jannaschii  
US-10-369-493-1061  
Query Match 7.6%; Score 115; DB 14; Length 1005;  
Best Local Similarity 19.4%; Pred. No. 0.9;  
Matches 59; Conservative 54; Mismatches 93; Indels 98; Gaps 11;  
Qy 7 EQTVVVVKSATADGALDPYNYKLDQVTPWKTDFE-----TIKELSRF-----KOEYSQ 56  
Db 605 DEILEDIKSGLNFKF--NFYNYLSAVSYLNSVDEGINRIKTEINIVSGWNKEKRE 661  
Qy 57 EASVL-----VGDIKVLMSODKYFEATQTVVWCGVVTOLLSAYILLFDEYNKKA 110  
Db 662 ELNKLREDEIREINRLKDKLNELNKKKE-----LIEIENRSKFKDYKEYLGL 710  
Qy 111 AOKDILIRILDGKVLNKAQKSLGSSQSFNNASGKLALD----- 152  
Db 711 TEKLEELKNIKDGLLEI-----YNICNSKILAIIDNIRKYNKEDIEIYLNK 757  
Qy 153 -----SQTNDFSEKSSYFQSQVDRIKEAVAGAAAGVAGPFGLLISYSIAAGVIRGKLI 208  
Db 758 ILEVNKEINDEIRISYINQKLDE-----INYNEE-----EKKI 792  
Qy 209 PEL-----NDRKAVONFTSLSVTVKQAKDIDAUKLATEIAAIGEIKTETETTFYV 264  
Db 793 KELYENKQELDNVREQKTEIETGIEYLVKQVESLKAFL-----KEMSNLEKEKLTKEV 848  
Qy 265 DYDD 268  
Db 849 EYLD 852  
RESULT 7  
US-10-369-493-5220  
; Sequence 5220, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 5220  
LENGTH: 2823  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-369-493-5220  
Query Match 7.5%; Score 112.5; DB 14; Length 2823;  
Best Local Similarity 19.9%; Pred. No. 6;  
Matches 73; Conservative 64; Mismatches 106; Indels 123; Gaps 16;  
Qy 6 ABOITVEV-----KSAIETADGALDPYNYKLDQVTPWKTDFETIKELSRFQKESQAS 59  
Db 1862 ABELLKAVTAQKLNIEIFEDLNKRIIDVLEQWMD-----YRETIYDVSKDTADAERMS 1915  
Qy 60 VLVG-----DIKVLMSODK-----YFEATQTV----- 83  
Db 1916 LVVGRINRYKEVSNIEKLRVEAEQAIYASRNSIEKARSEELMNMFEDEKINMTLAEL 1975  
Qy 84 ---YEWCGVVTOLLSAYILLFDEYNK-----KASAQKDILIRILDGKVLNKL 128  
Db 1976 PDLVEQCONITLL-----YSQLIDEYDEYVQTAGHAEKLEVQAQK-IVDRFVDTRTETEN 2031  
Qy 129 -----BAOKS-----LLGSSOSFNNASOKLALDSQLTNDSEKSSYFQSQVDRI 173  
Db 2032 PLKASHAYENIVREALKWATEAVDSAAEASAVSKMLGSESGESGDANEE--LRSQLEKL 2089  
Qy 174 RKEAYAGAAAGVAGPFGLLISYSIAAGVIE--GKLIPELNDRLKAVONFTSLSVTVKQ 231  
Db 2090 KNESSLSN-----VDNSNAVIVVEELKEKKDITDLRGLHNLKTSI----- 2131  
Qy 232 ANKDIDAUKLATEIAAIGEIKTETETTFYVDYDDLMLSLKGAQAKM-INTCNEYQQ 290  
Db 2132 -----VKRLGVIRNEASS-----WDDKDRMHSILKNGAKTAHERSANYKKE 2173  
Qy 291 RHGKKT 296  
Db 2174 SEGIKT 2179  
RESULT 8  
US-10-369-493-5221  
; Sequence 5221, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 5221  
; LENGTH: 2823  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-5221

Query Match 7.5%; Score 112.5; DB 14; Length 2823;  
Best Local Similarity 19.9%; Pred. No. 6;  
Matches 73; Conservative 64; Mismatches 106; Indels 123; Gaps 16;

Qy 6 ABOQTEVV-----KSAIETADGALDFYKYLDOVIPWKTDFEIKELSRFKQBYSQEAS 59  
Db 1862 ABEILKMVTAQKJNETIFEDLKNRIDLVEQWMD-----YRETIYDVSKKDTADAERMS 1915  
Qy 60 VLVG-----DIKVLMDSDK-----YFEATQTV----- 83  
Db 1916 LVVGRINRYKEVSEIEKLVEAEADQIAYSNSIEKARSEELMNMFEDEKINMTLAE 1975  
Qy 84 ---YEWCGVVTQLLSAYILLFDEYNEK-----KASAKQDLIRILDGYNKLN 128  
Db 1976 PDLVEQCNITLL---YSLIDYDEEYVQTAGRHAKELEVOAQK-IVDRFVDTRTETEN 2031  
Qy 129 -----EAKS---LLGSSQSFNNASGKLLALDSQTLWDFSEKSSYFQSOVDRI 173  
Db 2032 PLKASHAYENIVEALKNATEAVDSAAEASAVSKMLGSESGDANBES---LRSQLEKL 2089  
Qy 174 RKEAYAGAAGIVAGPFGLLIISYSIAAGVIE--GKLIPELNDRLKAVQNFPTSLSVTVKQ 231  
Db 2090 KNESSLSN-----VDNSNAVKIIVEELKEKKDLTRLCHLNLKTSI----- 2131  
Qy 232 ANKOIDAAKLKATEIAAIGBIKTETETTRFYVDYDDLMLSLKGAAKKM-INTCNEYQQ 290  
Db 2132 -----VKRLGVKNEASS---WDDKHDRMHILKNGAKTAKHARSANVKKE 2173  
Qy 291 RHGKKT 296  
Db 2174 SEGKKT 2179

RESULT 9  
US-10-369-493-6440  
; Sequence 6440, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6440  
; LENGTH: 1295  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-6440

Query Match 7.4%; Score 111; DB 14; Length 1295;  
Best Local Similarity 21.0%; Pred. No. 2.8;  
Matches 65; Conservative 58; Mismatches 93; Indels 94; Gaps 13;

Qy 6 ABOQTEVVKSAIETADGALDFYKYLDOVIPWKTDFET-----IKELSRFKQBYSQEASVL 61  
Db 710 ABETNEKLSRLASSE-----EQILDKNQOESLIDDLKEKLHSAESTNQELQVS 759  
Qy 62 VGDIKVLMDSDQKQFEATQTVYECGVVTTQLLSAYILLFDEYNEKKASA-----QKDI 115  
Db 760 LEMLKIEVSNAQKQWESE-----VLKESFEALQELSAQSVRSRVVDAVQEKDG 811  
Qy 116 LIRILDDGVNKLNEAKS---LLGSS-----OSFN---NASQK- 147  
Db 812 LLRLVDTLKLKIEDTEKSAQDLQSSVBEIKQLQLDQNFQKNAEVLSEKLNSHSHR 871

Qy 148 -LLALDSQTLNDFSEKSSYFQSQVDRIKRAYAGAAAGIVAGPFGLLIISYSIAAGVIEGK 206  
Db 872 DMVALASQL-EELQHLVWGESQVENVKEEL-----IGAKIMNKE 910  
Qy 207 LIPELN---DRLKAVQNFPTSLSVT-----VKQANKDIDAAKLKATEIAAIG 251  
Db 911 MVDELNAKLGDALLEGMBELKKSEVSEAKVQRREBELIAQVSKHRDQOE-QLQTLDELK 969  
Qy 252 EIKTETETTR 261  
Db 970 SAQSTETSR 979

RESULT 10  
US-10-424-599-212070  
; Sequence 212070, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 212070  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATUE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_33525C.1.pap  
US-10-424-599-212070

Query Match 7.3%; Score 110.5; DB 15; Length 564;  
Best Local Similarity 18.9%; Pred. No. 0.99;  
Matches 57; Conservative 59; Mismatches 115; Indels 71; Gaps 9;

Qy 11 EVKSAIETADGALDFYKYLDOVIPWKTDFETIKELSRFKQBYSQEASVLVGP---IKV 67  
Db 61 EVEREKURLVAEANLE-----KQAMDWMLAQEELKRLGEDAARHABESSETLDFRRVKK 114  
Qy 68 LLMDSDQKQFEATQTVYECGVVTTQLLSAYILLFDEYNEKKASQAQ-----KDILIRILD 121  
Db 115 LLNDVRSSELVSSQALASSRSKMEEQERLLELQSELGEQASVMSYMNKDAQIEV-- 172  
Qy 122 DGVNKLNEAQKSLGSSQSFNNASGKLLALDSQTLNDFSEKSSYFQSQVDRIKRAYAGA 181  
Db 173 -----ESERTKLRVAESRNRLELDLMEKELISELEELKERTSLEQAVKE----- 220  
Qy 182 AAGIVAGPFGLLIISYSIAAGVIEGKLIPELNDRLKAVQNFPTSLSVTVKQANKDIDA 241  
Db 221 -----VALQSELEKKTAEFRETSAVLQVKESLVDLAKL 254  
Qy 242 KLAETIAAIGBIKTETETTRFYVDYDDLMLSLKGAAKKMINTCNE---YQOR---HGKKT 296  
Db 255 E-----IQELKSEKASLQGLEEKLELS-----SARKMLGDVNOEIYDLKMLMHSKET 303  
Qy 297 LL 298  
Db 304 QL 305

RESULT 11  
US-10-282-122A-70920  
; Sequence 70920, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl  
APPLICANT: Hasebeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 70920  
LENGTH: 1189  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-10-282-122A-70920

Query Match 7.3%; Score 110; DB 15; Length 1189;  
Best Local Similarity 18.3%; Pred. No. 3;  
Matches 61; Conservative 60; Mismatches 130; Indels 82; Gaps 9;  
QY 7 EOTVEVVK-----SAIETADGALDFYKYLDOVIPWKTFTETIKELSRFKQYSEASV 60  
DB 163 EESAGVLKYKKAESIQKLDHTEDNLNRVEDILYDLGRVEPLKEEAIAKEYKQLSKE 222  
QY 61 LVGDIKVLMDSDQKYFEATQVYEWG-----GVVTOLLGAYILLFDEYNE 106  
DB 223 MEQSDVIVTSDIDHYTDNQRLDERNLHLSQQAEEGQQAQINQLQRY----- 273  
QY 107 KKAQAQKDLIRILDDGVNKLNEAQSILGSSQSFNNASGKLALDSQTLND-----FS 160  
DB 274 -KGRQQN-----DYDIEKLN---YELVKATENVEQLSGKLNVLBERKKNQSETNARYE 323  
QY 161 EKSSYFQSDVRIRKEAVAGAAAGIVAGPGLLIISYIAAGVIEGKLIPELNDRL----K 216  
DB 324 EELDNLESQDSIKNEKAQNE-----KLIALDLKNKQKQNLN 359  
QY 217 AVQNFETSLVTVKQANKDIDAALKLATEIAAIGEIKTETETTRFYVDYDMLSLK 276  
DB 360 EYQELLESLLYVSDQDEKLEIKNSYITLMSEQSDVNDIRFLEHTINENEAKSRLDS 419  
QY 277 AAKKMINTCNEYQQ-----RHGKKTLLV 300  
DB 420 RLVEAFNLQKDIQQNITQTNKAYQSSKKSMOKV 452

RESULT 12

US-10-028-248A-106  
Sequence 106, Application US/10028248A  
Publication No. US20030235882A1  
GENERAL INFORMATION:  
APPLICANT: Shimkets, Richard  
APPLICANT: Patturajan, Meera  
APPLICANT: Vernet, Corine  
APPLICANT: Casman, Stacie  
APPLICANT: Malyankar, Uriel  
APPLICANT: Shenoy, Suresh  
APPLICANT: Spytek, Kimberly  
APPLICANT: Gangolli, Esha  
APPLICANT: Miller, Charles  
APPLICANT: Boldog, Ferenc  
APPLICANT: Li, Li  
APPLICANT: Taupier Jr, Raymond J  
APPLICANT: Kekuda, Ramesh  
APPLICANT: Smithson, Glemnda  
APPLICANT: Zerhusen, Bryan  
APPLICANT: Liu, Xiaohong  
APPLICANT: Colman, Steven  
APPLICANT: Tchervnev, Velizar  
APPLICANT: Si, Jingsheng  
APPLICANT: Edinger, Shlomit  
APPLICANT: Stone, David  
APPLICANT: Sciore, Paul  
APPLICANT: Millet, Isabelle  
APPLICANT: Rothenberg, Mark  
TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods  
TITLE OF INVENTION: Thereof  
FILE REFERENCE: 21402-222  
CURRENT APPLICATION NUMBER: US/10/028,248A  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/256619  
PRIOR FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 60/262959  
PRIOR FILING DATE: 2001-01-19  
PRIOR APPLICATION NUMBER: 60/272408  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: 60/285189  
PRIOR FILING DATE: 2001-04-20  
PRIOR APPLICATION NUMBER: 60/308039  
PRIOR FILING DATE: 2001-07-26  
PRIOR APPLICATION NUMBER: 60/311266  
PRIOR FILING DATE: 2001-08-09  
NUMBER OF SEQ ID NOS: 211  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 106  
LENGTH: 1959  
TYPE: PRT  
ORGANISM: Gallus gallus  
US-10-028-248A-106

Query Match 7.2%; Score 108.5; DB 14; Length 1959;  
Best Local Similarity 21.0%; Pred. No. 8.1;  
Matches 66; Conservative 43; Mismatches 107; Indels 99; Gaps 11;  
QY 7 EOTVEVVKSAE---TADGALDFYKYLDOVIPWKTFTETIKE-----LSRFKQEYS 55  
DB 1166 EOEVTVLKKTLEDEAKTHEAOIQEMRQKHSQAI--EELAEQLEQTKRVKANLEKAKQALE 1223  
QY 56 QASVLVGDIDKVLMDSDQKYFEATQVYEWGVTQLLSAYILLFDEYNEKASAKQDI 115  
DB 1224 SERAELSNEVKVLLQKGDA-----EHRKRKVDQAQLOE 1256  
QY 116 L-----IRILDDGVNKLNEAQSILGSSQSFNNASGKLALDS---QLTNDFSEKSSY 165  
DB 1257 LQVKFTEGVRVKTALAEARNVKNLQVEL-----DNVTGLNQSDSKSIKAKDFSALSEQ 1309  
QY 166 FQSOVDRIKREAYAGAAAGIVAGPGLLIISYIAAGVIEGK---LIPELNDRLKAVQNF 222  
DB 1310 LQDTQELLQEEY-----RLKLSFSTKLQKQTEDEKNALKEQLSEEEAKRNLE 1356





ORGANISM: Rattus norvegicus  
US-10-028-248A-107  
Query Match 7.2%; Score 108.5; DB 14; Length 1999;  
Best Local Similarity 19.5%; Pred. No. 8.3; Indels 95; Gaps 13;  
Matches 66; Conservative 53; Mismatches 124; Indels 95; Gaps 13;  
QY 7 EQTVVVKSAIE---TADGALDFYNYKLDQVIP--WKTFTET---IKELSRFKQYSGE 57  
Db 1165 EQVNLKKTLEBEAKTHEAQIQEMRQKHSQAVEELAEQLEQTKRKVANLEKAKQTLNE 1224  
QY 58 ASVLVGDIKVLLMDSQDKYFEATQTVYWGCVVTTQLLSAYILLDFEYNEKKASAKDILI 117  
Db 1225 RGLANEVKKVLLQGRD-----SEHKRKVEAQLOELQ 1257  
QY 118 RILDGVNKLNEAQKSLGSSQSFNNASGKLLALD---SOLTNDFS----- 160  
Db 1258 VKFNEGERRVTELADKVKTKLOVELDNVTGLLSQSDSKSKLTQDFALESQLODQTOELLQ 1317  
QY 161 -----EKSYFQSVQDRIRKEAYAGAAAGIVAGPFGLLISYSIAAGV 202  
Db 1318 EENRQKLSLSTKLQVDEKNS-FREQLSEEEEAHNLEKQIA-----TLHAQV 1366  
QY 203 IEGKLIPELNDRKAVQNFSTLSVTVKQANKDIDAANKLKLATEIAAIGBIKTETETRF 262  
Db 1367 ADMK--KKMEDSVGCLL-----TAEVVKRKLQKLEGLSQRHEEKVAAYD--KLEKTKTRL 1418  
QY 263 YVDYDDLMLSLKGAAKKMTNCTN-EYQQRHGKKTLLLE 299  
Db 1419 QOELDDLLVDL-----DHQRSACNLEKKQKQKFDQLLAE 1452  
RESULT 16  
US-10-437-963-191043  
Sequence 191043, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 191043  
LENGTH: 815  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_8739C.1.pap  
US-10-437-963-191043  
Query Match 7.2%; Score 108; DB 16; Length 815;

ORGANISM: Rattus norvegicus  
US-10-028-248A-107  
Query Match 7.2%; Score 108.5; DB 14; Length 1999;  
Best Local Similarity 19.5%; Pred. No. 8.3; Indels 95; Gaps 13;  
Matches 66; Conservative 53; Mismatches 124; Indels 95; Gaps 13;  
QY 7 EQTVVVKSAIE---TADGALDFYNYKLDQVIP--WKTFTET---IKELSRFKQYSGE 57  
Db 1165 EQVNLKKTLEBEAKTHEAQIQEMRQKHSQAVEELAEQLEQTKRKVANLEKAKQTLNE 1224  
QY 58 ASVLVGDIKVLLMDSQDKYFEATQTVYWGCVVTTQLLSAYILLDFEYNEKKASAKDILI 117  
Db 1225 RGLANEVKKVLLQGRD-----SEHKRKVEAQLOELQ 1257  
QY 118 RILDGVNKLNEAQKSLGSSQSFNNASGKLLALD---SOLTNDFS----- 160  
Db 1258 VKFNEGERRVTELADKVKTKLOVELDNVTGLLSQSDSKSKLTQDFALESQLODQTOELLQ 1317  
QY 161 -----EKSYFQSVQDRIRKEAYAGAAAGIVAGPFGLLISYSIAAGV 202  
Db 1318 EENRQKLSLSTKLQVDEKNS-FREQLSEEEEAHNLEKQIA-----TLHAQV 1366  
QY 203 IEGKLIPELNDRKAVQNFSTLSVTVKQANKDIDAANKLKLATEIAAIGBIKTETETRF 262  
Db 1367 ADMK--KKMEDSVGCLL-----TAEVVKRKLQKLEGLSQRHEEKVAAYD--KLEKTKTRL 1418  
QY 263 YVDYDDLMLSLKGAAKKMTNCTN-EYQQRHGKKTLLLE 299  
Db 1419 QOELDDLLVDL-----DHQRSACNLEKKQKQKFDQLLAE 1452  
RESULT 15  
US-10-107-782-107  
Sequence 107, Application US/10107782  
Publication No. US20040018970A1  
GENERAL INFORMATION:  
APPLICANT: Boldog, Ferenc,  
APPLICANT: Casman, Stacie  
APPLICANT: Colman, Steve,  
APPLICANT: Edinger, Shlomit,  
APPLICANT: Gangolli, Esha,  
APPLICANT: Kekuda, Rameesh,  
APPLICANT: Li, Li,  
APPLICANT: Liu, Xiaohong,  
APPLICANT: Malyankar, Uriel,  
APPLICANT: Miller, Charles,  
APPLICANT: Millet, Isabelle,  
APPLICANT: Patturajan, Meera,  
APPLICANT: Rothenberg, Mark,  
APPLICANT: Sciore, Paul,  
APPLICANT: Shenoy, Suresh,  
APPLICANT: Shimkets, Richard,  
APPLICANT: Si, Jingsheng,  
APPLICANT: Smithson, Glennda,  
APPLICANT: Spvtek, Kimberly,  
APPLICANT: Stone, David,  
APPLICANT: Taupier, Raymond, jr.,  
APPLICANT: Tchernev, Velizar,  
APPLICANT: Vernet, Corine,  
APPLICANT: Zerhusen, Brian  
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF  
FILE REFERENCE: 21402-222CIP  
CURRENT APPLICATION NUMBER: US/10/107,782  
CURRENT FILING DATE: 2002-03-27  
PRIOR APPLICATION NUMBER: 10/028,248  
PRIOR FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/256,619  
PRIOR FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 60/262,959  
PRIOR FILING DATE: 2001-01-19  
PRIOR APPLICATION NUMBER: 60/272,408  
PRIOR FILING DATE: 2001-02-28

Best Local Similarity 20.7%; Pred. No. 2.7;  
Matches 62; Conservative 62; Mismatches 95; Indels 80; Gaps 14;

QY 38 KTFEDTIK-ELSRFKQESQ--EASVL-----VGDIKVLLMDSQDKYFEATQTVVEW 86  
DB 290 KTLAEELREHRLLESSLAQAGDDKVLKAKELKGLGVNVL-----QEKVLLSSEI-DN 344  
QY 87 CGV-----VTQLLSA-----YILLFDEYNKKASQAQ-KDILIRILDGVNKNKNEAQKSLGS 137  
DB 345 KGIIRRELSLLSSKEADYRNLCSFDQTKESLELAIAKIQQLEBEVHTRNDLSSKISS 404  
QY 138 -----SOSFNASGKLLALDSQTNDFSEKSSYFQSQVDRIKREAYAGAAAGIVAGPF 190  
DB 405 IDLNEELQALNSAKNEAEKLSLTQYDTLKASSEARENSR-----449  
QY 191 GLIISYSIAAGVIEGKLIPELNDRLKAVQNFTSLSVTKQANKDIDAACKLKLATEIAAI 250  
DB 450 -LLEKONWIKQLGKLSALSSDKREN-----IAALNKELDATKAMLENEVAV 500  
QY 251 GEIK-----TETETFRFYVDYD-----DLMLSLKGAAKMINTCNEYQQ 290  
DB 501 KSLRESLQSTEEALDTSRSEVSKLSVELDEANRNMQDLVLI-----SKLQDEFNMQE 554

RESULT 17  
US-10-369-493-3279  
; Sequence 3279, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052) B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 3279  
; LENGTH: 1965  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
US-10-369-493-3279

Query Match 7.2%; Score 108; DB 14; Length 1965;  
Best Local Similarity 21.3%; Pred. No. 8.9;  
Matches 64; Conservative 47; Mismatches 113; Indels 76; Gaps 12;

QY 6 AEQTEVVKSAIETADGALDFYKYLDOVIPKTFD-----ETIKELSRFKQESQ 56  
DB 57 AKYREALKQELHAHRT-----NDWLDNELTKNAEAIKFRKEKGARIAEQLRMEDANS 112  
QY 57 EASVLVGDIKVL---LMDSDQKYFEATQTVVEWC-----GVVTQLLSA---YILLFDE 103  
DB 113 TIESITREQVLRKLEQAQDAEETLTKVQQLQEAARTTEGFKQELSAKRLVELKQ 172  
QY 104 YNEKASAKQITLIR---ILDDGVNKNKNEAQKSLGSQSOSFNASGKLLALDSQTNDFS 160  
DB 173 QSETHRNRLKEVELRLQIKDDHANEIRIRRELEKE-----DHAQT---E 217  
QY 161 EKSSYFQSQVDRIKE-----AYAGAAAGIVAGPGLIISISIAAG 201  
DB 218 QRAQELQNEVDRIKASSDLRGSPGTPQTPKGDGSAFATRAG---SPFGTPLS-----268  
QY 202 VIEGKLIPELNDRLKAVQNFTSLSVTKQANK---DIDAACKLKLATEIAAIGETKETE 258  
DB 269 -IRGRAGQATDALBELYNVKQLAGEKRCCKLQEBELDDAVAMLEAKMPEIDELNASE 327

RESULT 18  
US-10-282-122A-53523  
; Sequence 53523, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 53523  
; LENGTH: 742  
; TYPE: PRT  
; ORGANISM: Clostridium difficile  
US-10-282-122A-53523

Query Match 7.1%; Score 107.5; DB 15; Length 742;  
Best Local Similarity 21.2%; Pred. No. 2.6;  
Matches 66; Conservative 49; Mismatches 134; Indels 63; Gaps 10;

QY 18 ETAGALDFYKYLDOVIPKTFDTEIKELSRFKQESQASVLVDGIKVLMDSDQKYF 77  
DB 236 KTVDLASDATSKLSDTVKDIKSDLPTIKT-----LNDTKLLSSDLKFLDFTNDL 288  
QY 78 EATQTVVEWCQVVTQLLSAYILLFDEYNKKASQAQDKILIRLD-----DGVNKNLEAQ 131  
DB 289 ELSPLIKSDMLMDVL-----SSSASSUTLNLIDAVNSGSEVDPKLI 333  
QY 132 KSLIGSSQSFNNSGKLLALDSQTN-----DFSEKSSYFQSQVDRIKREAYA 179  
DB 334 SEKLSNLQSLNDTLVDFTLKNQLTSNNRLDDVDINLEDSSNKIDSSISTLNDIKNKVIS 393  
QY 180 GAAGIVAGPGLIISYSIAAGVIEGKLIPELNDRL-KAVQNFTSLSVTKQANKDI-- 236  
DB 394 GQPSISA--LNNVLSLNSGIGRINLNLNNFDSKISKIPINNIFAN---SIKVANDIITV 448  
QY 237 -DAKLLK-----LATEIAAIGETKETEITFRFYVDYDDLMLSLKGAAKMINTCNEY 288  
DB 449 LDKAEAKLPKVEILTTSLKSGNAQSILIR-----ERLPLAKGLMDLIDILSKI 501

SEQ ID NO 5109  
LENGTH: 1938  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-369-493-5109

Query Match 7.1%; Score 106.5; DB 14; Length 1938;  
Best Local Similarity 17.3%; Pred. No. 12;  
Matches 52; Conservative 63; Mismatches 112; Indels 73; Gaps 8;

QY 5 FAEQTVVVK-----SAIETADGALDFYKYLQVVPWKTFDETIKELSRFK 51  
DB 1315 FSSQLVEAKAAAEDELHERQEFHAAACKNLEHLEDOCHELLEEQNGK--DDIQQLSRIN 1372  
QY 52 QEYSQ-----EASVLVG--DIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEY 104  
DB 1373 SEISQWKARYEGEGLVSGEELKQKQNRVMDLQEAALSAQKVKVLSLEKAKGLLAET 1432  
QY 105 NEKKASAKDILI-----RIIDDGVNKLNEAKSKLLSSQSFNNASGKLLALD 152  
DB 1433 EDARSDVDRHLTVIASLEKKQKQAFDKIVDDMKRVDIOKEIDATTRDSRNTSTEVFKLR 1492  
QY 153 SOLTNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELN 212  
DB 1493 SSMN-----LSEQIETLR-----NKIFSQBIRDIN 1520  
QY 213 DRLKAVQNFTSLSVTVKQAKD-----IDAANKLATEIAAIGEIKTETETTRFYVD 265  
DB 1521 EQITQGGRTYQEVHKSVRLEQEKDLOHALDEAAEALEAESKVLRLQIEVQOIRSEIE 1580

RESULT 21  
US-10-369-493-5110  
Sequence 5110, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
PRIOR FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 5110  
LENGTH: 1938  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-369-493-5110

Query Match 7.1%; Score 106.5; DB 14; Length 1938;  
Best Local Similarity 17.3%; Pred. No. 12;  
Matches 52; Conservative 63; Mismatches 112; Indels 73; Gaps 8;

QY 5 FAEQTVVVK-----SAIETADGALDFYKYLQVVPWKTFDETIKELSRFK 51  
DB 1315 FSSQLVEAKAAAEDELHERQEFHAAACKNLEHLEDOCHELLEEQNGK--DDIQQLSRIN 1372  
QY 52 QEYSQ-----EASVLVG--DIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEY 104  
DB 1373 SEISQWKARYEGEGLVSGEELKQKQNRVMDLQEAALSAQKVKVLSLEKAKGLLAET 1432  
QY 105 NEKKASAKDILI-----RIIDDGVNKLNEAKSKLLSSQSFNNASGKLLALD 152  
DB 1433 EDARSDVDRHLTVIASLEKKQKQAFDKIVDDMKRVDIOKEIDATTRDSRNTSTEVFKLR 1492  
QY 153 SOLTNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELN 212

QY 289 QQRHGKKTLLV 300  
DB 502 SNGEDMKLVSL 513

RESULT 19  
US-10-369-493-9374  
Sequence 9374, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 9374  
LENGTH: 857  
TYPE: PRT  
ORGANISM: Xylella fastidiosa  
US-10-369-493-9374

Query Match 7.1%; Score 107.5; DB 14; Length 857;  
Best Local Similarity 23.9%; Pred. No. 3.2; Indels 104; Gaps 14;  
Matches 68; Conservative 45; Mismatches 104; Indels 67; Gaps 14;

QY 7 EQTVVVKSAIETADGALDFYK-----YLDQVVPWKTFDETIKELSRFKQEYSQ 60  
DB 574 EEAIVSDAVRSRTGLSDPNRPSGFLFLGPTGVGKT--ELCKALAEFLD-SQDAMV 630  
QY 61 LVGDIKVLMDSDQKYFEAT-----QTVYEWCGVVTQLLSA---YILLFDEYNEKKAS 110  
DB 631 -----RIDMSEFEKHSVARLIGAPPYGYEGGYTELVRPYSLLILDEV-EKAS 684  
QY 111 AQDILRIIDDGVNKLNEAKSKLLSSQSFNNASGKLLALDLSQTLNDFSEKSSYFQSQV 170  
DB 685 DVNILLQVLDG--RLTDCQ-----GRTVDFRNT--VIVMTSLGSHQIQELSGDDS-- 733  
QY 171 DRIKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELNDRKAVQNFTSLSVTVK 230  
DB 734 ---PEVVTOMKAAM-----GVQAHPRPFINRLDDIVVF----- 766  
QY 231 QANKDIDAANKLATEIAAIGEIKTETETTRFYVDYDMLSL 274  
DB 767 ---HPLDKAQIKQIARIQLRGLERLAE--SELKLLDLDRALELL 806

Query Match 7.1%; Score 107.5; DB 14; Length 857;  
Best Local Similarity 23.9%; Pred. No. 3.2; Indels 104; Gaps 14;  
Matches 68; Conservative 45; Mismatches 104; Indels 67; Gaps 14;

QY 7 EQTVVVKSAIETADGALDFYK-----YLDQVVPWKTFDETIKELSRFKQEYSQ 60  
DB 574 EEAIVSDAVRSRTGLSDPNRPSGFLFLGPTGVGKT--ELCKALAEFLD-SQDAMV 630  
QY 61 LVGDIKVLMDSDQKYFEAT-----QTVYEWCGVVTQLLSA---YILLFDEYNEKKAS 110  
DB 631 -----RIDMSEFEKHSVARLIGAPPYGYEGGYTELVRPYSLLILDEV-EKAS 684  
QY 111 AQDILRIIDDGVNKLNEAKSKLLSSQSFNNASGKLLALDLSQTLNDFSEKSSYFQSQV 170  
DB 685 DVNILLQVLDG--RLTDCQ-----GRTVDFRNT--VIVMTSLGSHQIQELSGDDS-- 733  
QY 171 DRIKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELNDRKAVQNFTSLSVTVK 230  
DB 734 ---PEVVTOMKAAM-----GVQAHPRPFINRLDDIVVF----- 766  
QY 231 QANKDIDAANKLATEIAAIGEIKTETETTRFYVDYDMLSL 274  
DB 767 ---HPLDKAQIKQIARIQLRGLERLAE--SELKLLDLDRALELL 806

RESULT 20  
US-10-369-493-5109  
Sequence 5109, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374

Query Match 7.1%; Score 106.5; DB 14; Length 1938;  
Best Local Similarity 17.3%; Pred. No. 12;  
Matches 52; Conservative 63; Mismatches 112; Indels 73; Gaps 8;

QY 5 FAEQTVVVK-----SAIETADGALDFYKYLQVVPWKTFDETIKELSRFK 51  
DB 1315 FSSQLVEAKAAAEDELHERQEFHAAACKNLEHLEDOCHELLEEQNGK--DDIQQLSRIN 1372  
QY 52 QEYSQ-----EASVLVG--DIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEY 104  
DB 1373 SEISQWKARYEGEGLVSGEELKQKQNRVMDLQEAALSAQKVKVLSLEKAKGLLAET 1432  
QY 105 NEKKASAKDILI-----RIIDDGVNKLNEAKSKLLSSQSFNNASGKLLALD 152  
DB 1433 EDARSDVDRHLTVIASLEKKQKQAFDKIVDDMKRVDIOKEIDATTRDSRNTSTEVFKLR 1492  
QY 153 SOLTNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELN 212

Db 1493 SSMON-----USEQIETURR-----NKIPSOEIRDIN 1520

Qy 213 DRLXAVQNFYSLSVTVKQANKD-----IDAAKLKLATETAAIGETETETTRFYVD 265

Db 1521 EQITOGGRTYQVBHKSVRRLQKDEQKDELQALDEAAEALESKVLRRLQIEVQQRSEIE 1580

RESULT 22

US-10-282-122A-71235

; Sequence 71235, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Hasebeck, Robert

; APPLICANT: Ohleen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-06-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 71235

; LENGTH: 3692

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-10-282-122A-71235

Query Match 7.1%; Score 106.5; DB 15; Length 3692;

Best Local Similarity 21.5%; Pred. No. 28;

Matches 71; Conservative 49; Mismatches 129; Indels 81; Gaps 12;

Qy 11 EVKSAIETADGALDFYKYLQVVPWKTFDTIKELSR-----FKQEYSQ-EA 58

Db 564 EQVKNKIIPSNYTLASYNKY-----NKLKERAQTVLDEETNTPFNQRYSTQI 612

Qy 59 SVLVGDIKVLMM-----DSQDKYFEATQTYEWCVVQTLLSAYILLFDEYNEKKASA 111

Db 613 DDLHELQTLNIRVRSASREINDKAQEMTDVAYDSTELTTEKDT---LVDQIENHNEI 669

Qy 112 QKDILIRILDGCVKLNKAQKSLGSS-----QSEFNASGKLLALDSQLTNDIFS 160

Db 670 SNNIDELTDDGVERVKAGLHTLESDDTTPHVPVKPNARQVNNRA-----DOQKTLIRN 723

Qy 161 EKSYFQSOVDRIK-EAYAGAAAGIVNGPGLIISIAAGVIEGKLIPELNDELKAVQ 219

Db 724 NHEATTBEQNEAIRQVEAHSSDA-----IAKIGEAETDTTWN-----ARD 764

Qy 220 NFFTSLSVTVKQANKDIDAAKCLKLATEIAAIGETETETTRFYVDYDDLMLSLKGAAK 279

Db 765 NGTKLIATDVNPYTKAEA---RAAVTNSANSKIKDINNNTQATLDERDAIALVNRSKD 821

Qy 280 KMINTCNEYQ-----QRHGKKTLLVVP 301

Db 822 EAIQNTAQNAGNDVDVTEAQNNGTNTTQQVP 851

RESULT 23

US-10-028-248A-105

; Sequence 105, Application US/10028248A

; Publication No. US20030235882A1

; GENERAL INFORMATION:

; APPLICANT: Shinkete, Richard

; APPLICANT: Patturajan, Meera

; APPLICANT: Vernet, Corine

; APPLICANT: Casman, Stacie

; APPLICANT: Malvankar, Uriel

; APPLICANT: Shenoy, Suresh

; APPLICANT: Spytek, Kimberly

; APPLICANT: Gangolli, Esha

; APPLICANT: Miller, Charles

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Smithson, Glennda

; APPLICANT: Zerhusen, Bryan

; APPLICANT: Liu, Xiaohong

; APPLICANT: Colman, Steven

; APPLICANT: Tchernev, Velizar

; APPLICANT: Si, Jingsheng

; APPLICANT: Edinger, Shlomit

; APPLICANT: Stone, David

; APPLICANT: Sciore, Paul

; APPLICANT: Millet, Isabelle

; APPLICANT: Rothenberg, Mark

; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods

; FILE REFERENCE: 21402-222

; CURRENT APPLICATION NUMBER: US/10/028,248A

; CURRENT FILING DATE: 2001-12-19

; PRIOR FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 60/256619

; PRIOR FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 60/262959

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/272408

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 60/285189

; PRIOR FILING DATE: 2001-04-20

; PRIOR APPLICATION NUMBER: 60/308039

; PRIOR FILING DATE: 2001-07-26

; PRIOR APPLICATION NUMBER: 60/311266

; PRIOR FILING DATE: 2001-08-09

; NUMBER OF SEQ ID NOS: 211

; SOFTWARE: PatentIn ver. 2.1

; SEQ ID NO 105

; LENGTH: 1961

; TYPE: PRT

; ORGANISM: Rattus norvegicus

US-10-028-248A-105

Query Match 7.0%; Score 106; DB 14; Length 1961;

Best Local Similarity 21.7%; Pred. No. 13;

Matches 68; Conservative 50; Mismatches 100; Indels 96; Gaps 15;

Qy 7 EQTVEVKSATETADGALDFYKYLQVVPWKTFDTIKELSRFKQEYSQEAASVLGVGIK 66



Db 175 -----AFNIESGVTTHASHM-NDQPVIDAY-HSDLRTR-----AASQSIIP-- 215  
Qy 190 FGLIISYSIAAGVIEGKLIPELNDRLKAVONFSTLSVTVKQAKDIDAOKLKLATEIAA 249  
Db 216 ----VDTKLAAGIT--RIFPKFSNFEALSVRVPINVTAILDSVTVSA-----AVNVSE 264  
Qy 250 IGEL--KTETETTRFVVDVDDLML 271  
Db 265 VNQLQKSAVSVFRGIVDFDTLPL 288

RESULT 26  
US-10-437-963-107270  
; Sequence 107270, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 107270  
; LENGTH: 737  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; PEPTIDE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_11638C.1.l.pap  
US-10-437-963-107270

Query Match 7.0%; Score 105; DB 16; Length 737;  
Best Local Similarity 21.3%; Pred. No. 4.3;  
Matches 69; Conservative 63; Mismatches 126; Indels 66; Gaps 12;  
Qy 4 IPAEQTVVVKSAIETAD---GALDFYKYLQVPIWKTFD---ETIKELSRFKQSYQ 56  
Db 249 LVAEQKLNICEAIEERLKMELGALTEANE-----AAAKAFDTQNEBITKELEDLTKLEE 303  
Qy 57 -----EASVLGDIKVLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKAS 110  
Db 304 IKTNKDLAESNGKLRSELLSSEKYSOSEAEV---KYLQVMGA-VVEAKEAAAKAFA 358  
Qy 111 AOKDILIRLDDGVNKLNEAOKSLGSSOSFNNAAGKLIALDSQLTNDPSEKSSYFQSQV 170  
Db 359 AEKEIDIMESDNLKRVKEIQDLSKLLVSENEDELASEILSM-----KQKHGQFEVEV 410  
Qy 171 DRIRKEAYAGAAGIVAGFGLIISYSTAAGVIEGKLIPELNDRLKAVONFSTS----- 224  
Db 411 TSLKKELGALAEAKSIT-----TKAFEVEKTEILKELEDLKRKVVIEIQTNKDLVEV 461  
Qy 225 -----LSVTVQANKDIDAOKLKLATEIAAIGEIKTETETTRFFVYDIDLMLSLK 275  
Db 462 GNDKLRDLVLSAQKQKQSTLEVEANNLKM--ELGALVEAK-EVATKAFDAEKAKI----- 512  
Qy 276 GAAKMMINTCNEYQQRHGKKTLE 299  
Db 513 --TKELEDVKKRQWERIQVKDLVE 534

RESULT 27  
US-09-815-242-10796  
; Sequence 10796, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10796  
; LENGTH: 722  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-10796

Query Match 6.9%; Score 104; DB 9; Length 722;  
Best Local Similarity 19.8%; Pred. No. 5;  
Matches 66; Conservative 51; Mismatches 109; Indels 108; Gaps 12;  
Qy 16 ALETADGALDFYKYLQVPIWKTFDETIKELSRFKQSY-----SQE 57  
Db 122 AVKGNKDIKSYDDDLKGTVAAKVGTESANFLEKNEKYDYTIKNFDDATGLYKALENGE 181  
Qy 58 ASVLGDIKVL-----LMDSQD-----KYFEATQTVYEWCG 88  
Db 182 ADAIVDDYPVLGYAVKNGQKLQVGDKETGSSGYFAVKGQNPelikfknaglkngkng 241  
Qy 89 VVTQLLSAYILLFDEYNEKASAKDILIRLDDGVNKLNEAOKSLGSSQS-----PN 142  
Db 242 TYDKILNNYLATGDETNTQDAGEQ-----MKKITPKKEYVIASDSTFAPFEFQ 290  
Qy 143 NASGKLALDSQLTNDPSEKSSY-----FQSOVDRIKEAYAGAAGIVAGFGLII 194  
Db 291 NAQGDYVIDVLDVKRAAELOQTFVEFKFIFGSSAVQAVE-----SGQADGWAG----- 340  
Qy 195 SYSTAAGVIEGKLIPELNDRLKA-----VONFSTLSVTVKQAN-----KQIDAAK-----LK 242  
Db 341 -----MTITDDRKKAEDFSVPYFDSGFIQIAVKGKNDIKSYDIDLKGGKVGK 387  
Qy 243 LATEIA-AIGIKTETETTRFFVYDIDLMLSLK 275  
Db 388 IGTESADFLEKNNKKYDYSIKYLTDTTDLALYSALE 421

RESULT 28  
US-09-841-260-139  
; Sequence 139, Application US/09841260  
; Publication No. US20030175700A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Probst, Peter  
; APPLICANT: Stromberg, Erika Jean  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS

us-09-993-292b-24.rapb

Wed Feb 2 11:26:41 2005

Db 268 LQAQNNSPDNIAATKELIDAAETKVNELKQHTGL-----TDSPLVKK 311

QY 159 FSEKSSYFQSOVDRIKEAVAGAAAGIVAGPGLIISYIAAGVIEGK-----LIPE 210

Db 312 AEEQISQAQKDIQEIKP---SGSDIPV-GPSSGAASAGSAGALKSSNNSGRISLLDD 367

QY 211 LNDRLKAV-----QNFFTSLSVTVKQAKDIDAOKLATE-----IAAIGIKTETETTR 261

Db 368 VDNEMAAIALQGFMSIEQFNNPATAKELQAMEAQLTAMSDQLVGDGELPAEIOAIK 427

QY 262 FYVDYDDLMLSLK 275

Db 428 -----DALAQALK 435

RESULT 30

US-10-762-058-139

; Sequence 139, Application US/10762058

; Publication No. US20040137007A1

; GENERAL INFORMATION:

; APPLICANT: Bhatia, Ajay

; APPLICANT: Probst, Peter

; APPLICANT: Stromberg, Erika Jean

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS

; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.515D1

; CURRENT APPLICATION NUMBER: US/10/762,058

; CURRENT FILING DATE: 2004-01-15

; NUMBER OF SEQ ID NOS: 140

; SEQ ID NO 139

; LENGTH: 660

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-10-762-058-139

Query Match 6.9%; Score 103.5; DB 16; Length 660;

Best Local Similarity 21.7%; Pred. No. 4.9;

Matches 68; Conservative 48; Mismatches 119; Indels 79; Gaps 14;

QY 8 QTEVVVKSATETADGALDFYKYL-----DQV-----IPWKTFF 40

Db 155 EVNNIKKALEAQKDTIDKLNKLVTLQNKSLTEVLKTTDSADQIPAINSOLEINKNSA 214

QY 41 DETIKELSRFKQYSEASVILGDIKVLMDSDQ--KYFEATQTVYEWGVVTVQLLSAYI 98

Db 215 DQIIKDLER--QNISYEA-VLTNAGEVIKASSEAGIKLQALQSIVD-AGDQSO---AAV 267

QY 99 LLFDEYNEKKASAKDILIRILDDGVNKLNEAKSLGSSQSFNNASGKLLALDSQLTND 158

Db 268 LQAQNNSPDNIAATKELIDAAETKVNELKQHTGL-----TDSPLVKK 311

QY 159 FSEKSSYFQSOVDRIKEAVAGAAAGIVAGPGLIISYIAAGVIEGK-----LIPE 210

Db 312 AEEQISQAQKDIQEIKP---SGSDIPV-GPSSGAASAGSAGALKSSNNSGRISLLDD 367

QY 211 LNDRLKAV-----QNFFTSLSVTVKQAKDIDAOKLATE-----IAAIGIKTETETTR 261

Db 368 VDNEMAAIALQGFMSIEQFNNPATAKELQAMEAQLTAMSDQLVGDGELPAEIOAIK 427

QY 262 FYVDYDDLMLSLK 275

Db 428 -----DALAQALK 435

RESULT 31

US-10-282-122A-58016

; Sequence 58016, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.515

; CURRENT APPLICATION NUMBER: US/09/841,260

; CURRENT FILING DATE: 2001-04-23

; NUMBER OF SEQ ID NOS: 140

; SEQ ID NO 139

; LENGTH: 660

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-09-841-260-139

Query Match 6.9%; Score 103.5; DB 10; Length 660;

Best Local Similarity 21.7%; Pred. No. 4.9;

Matches 68; Conservative 48; Mismatches 119; Indels 79; Gaps 14;

QY 8 QTEVVVKSATETADGALDFYKYL-----DQV-----IPWKTFF 40

Db 155 EVNNIKKALEAQKDTIDKLNKLVTLQNKSLTEVLKTTDSADQIPAINSOLEINKNSA 214

QY 41 DETIKELSRFKQYSEASVILGDIKVLMDSDQ--KYFEATQTVYEWGVVTVQLLSAYI 98

Db 215 DQIIKDLER--QNISYEA-VLTNAGEVIKASSEAGIKLQALQSIVD-AGDQSO---AAV 267

QY 99 LLFDEYNEKKASAKDILIRILDDGVNKLNEAKSLGSSQSFNNASGKLLALDSQLTND 158

Db 268 LQAQNNSPDNIAATKELIDAAETKVNELKQHTGL-----TDSPLVKK 311

QY 159 FSEKSSYFQSOVDRIKEAVAGAAAGIVAGPGLIISYIAAGVIEGK-----LIPE 210

Db 312 AEEQISQAQKDIQEIKP---SGSDIPV-GPSSGAASAGSAGALKSSNNSGRISLLDD 367

QY 211 LNDRLKAV-----QNFFTSLSVTVKQAKDIDAOKLATE-----IAAIGIKTETETTR 261

Db 368 VDNEMAAIALQGFMSIEQFNNPATAKELQAMEAQLTAMSDQLVGDGELPAEIOAIK 427

QY 262 FYVDYDDLMLSLK 275

Db 428 -----DALAQALK 435

RESULT 29

US-10-007-693-139

; Sequence 139, Application US/10007693

; Publication No. US20020146776A1

; GENERAL INFORMATION:

; APPLICANT: Bhatia, Ajay

; APPLICANT: Probst, Peter

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT

; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.515C2

; CURRENT APPLICATION NUMBER: US/10/007,693

; CURRENT FILING DATE: 2001-12-05

; NUMBER OF SEQ ID NOS: 157

; SEQ ID NO 139

; LENGTH: 660

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-10-007-693-139

Query Match 6.9%; Score 103.5; DB 13; Length 660;

Best Local Similarity 21.7%; Pred. No. 4.9;

Matches 68; Conservative 48; Mismatches 119; Indels 79; Gaps 14;

QY 8 QTEVVVKSATETADGALDFYKYL-----DQV-----IPWKTFF 40

Db 155 EVNNIKKALEAQKDTIDKLNKLVTLQNKSLTEVLKTTDSADQIPAINSOLEINKNSA 214

QY 41 DETIKELSRFKQYSEASVILGDIKVLMDSDQ--KYFEATQTVYEWGVVTVQLLSAYI 98

Db 215 DQIIKDLER--QNISYEA-VLTNAGEVIKASSEAGIKLQALQSIVD-AGDQSO---AAV 267

QY 99 LLFDEYNEKKASAKDILIRILDDGVNKLNEAKSLGSSQSFNNASGKLLALDSQLTND 158



APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

## TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 58016

LENGTH: 1020

TYPE: PRT

ORGANISM: Enterococcus faecium

US-10-282-122A-58016

Query Match 6.9%; Score 103.5; DB 15; Length 1020;

Best Local Similarity 19.8%; Pred. No. 8.9;

Matches 65; Conservative 41; Mismatches 90; Indels 133; Gaps 12;

Qy 3 GIFAQTVVVK-----SAETADG---ALDFYKYLDOVVPKTFDETIELSR 49

Db 159 GIF-EAAGVLKYKORKKAEQKLFETEDNLSRVQDIIHELEEQUTPLAAQSEAAKEFLR 217

Qy 50 PQEYSQ-EASVLVGDIKVLLMDSQDKYFEATQTVYWGCVVTVQLLSAYILLFDEYNEKK 108

Db 218 LKETLTQDVSLMVAEIKTKD-----WDNKQ 245

Qy 109 ASAQKDIILRILDDGVNKLNEA---QKSLGSSQSFNNAAGKLLALDSQTNDFPEKSSY 165

Db 246 AQLAKFNL-----ELGKLSIESIQEESILAKORKENAQADRLIEKNOQVLLDLSEKLQ 299

Qy 166 FQSOVDRIKRAYAGAAAGIAGVAGPGGLIISYIAGVIEGKLIPELNDRLKAVQNFFTSL 225

Db 300 TEGOKD-----VLQERTKHTQKSOEY 321

Qy 226 SVTVQANKNDIAAKLKLATEIAAIGEIKETETETTRFVYVDVDDLMLSLKGAARK----- 280

Db 322 QTSLAEAK-----KVK-----HFEKLQESLMKAAAEKETEIQ 354

Qy 281 -----MINTCNEYQ--QRHGKTKLLEVPD 302

Db 355 KAEANLINTQOELEKYQKSTKELLAELRD 383

RESULT 32

US-10-282-122A-51864

; Sequence 51864, Application US/10282122A

Publication No. US20040029129A1

## GENERAL INFORMATION:

APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

## TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 51864

LENGTH: 1163

TYPE: PRT

ORGANISM: Clostridium acetobutylicum

US-10-282-122A-51864

Query Match 6.9%; Score 103.5; DB 15; Length 1163;

Best Local Similarity 21.7%; Pred. No. 11;

Matches 81; Conservative 55; Mismatches 142; Indels 95; Gaps 14;

Qy 1 MTGIFAQTVVVKSAIETADG---ALDFYKYLDOVVPKTFDETIK-----ELSL 48

Db 112 ITGDEVEVLGSAKSVNEKQOEIIGLSLDDFTRTV--VLPOGKFSEFLKLEGKERNMLE 169

Qy 49 RF--KQEYSQSEAS-----VLVGDIKVLLMDSQDKYFEATQTVYWGCVVT 91

Db 170 RLFLNQLQYGDLSFPLARKIRKEREKENVLGGELKGVENINEDVLKERELLKENDDFN 229

Qy 92 QLLSAYILLFPEYNEKKAQAKDILI-----RILDDGVNKLNE-----AQKSLIGSSQS--- 140

Db 230 EASKEYLKAABEYNEKGVWGLQIEIEKNRVRKDLMEKDEIDLKEKRALGESSKVK 289

Qy 141 -----FNNAASKLALDLSQLTN-----DFSEKSSYFQSQVDRIKRAYAGAA 182

Db 290 PYINVENTLKQIDILKEQILSRENTWKAISLEKMEKKLSIAKDNEK----- 339

Qy 183 AGIVAGPFGLIISYIAGVIEGKLIPELNDRLKAVQNFFTSLV-----TVQAKND 235

Db 340 ----ALPKFMKHHIILDAIKEKDLDDNIKLEKKRLQCKIEKLSLEASKEELIKQIKD 395

Qy 236 IDAAKLKLATEIAAIGEIKTETETTRFVYVDVDDLMLSLKGAAKMIN-----TCNEYQ 289

396 IDSLTLKIQNLESKIDNLKVPEE-----YKKNKINEGIFLLRNYDEKLKHKNNKGLGDCDKFO 451  
290 ----QRHGKTKLL 298  
452 VDFEAKSKKEML 464

Db 396 IDSLTLKIQNLESKIDNLKVPEE-----YKKNKINEGIFLLRNYDEKLKHKNNKGLGDCDKFO 451  
QY 290 ----QRHGKTKLL 298  
Db 452 VDFEAKSKKEML 464

RESULT 33  
US-10-408-765A-741  
; Sequence 741, Application US/10408765A  
; Publication No. US200401018741  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
; FILE REFERENCE: 660088.465  
; CURRENT APPLICATION NUMBER: US/10/408,765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 741  
; LENGTH: 5373  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-408-765A-741

Query Match 6.9%; Score 103.5; DB 16; Length 5373;  
Best Local Similarity 23.0%; Pred. No. 86;  
Matches 56; Conservative 49; Mismatches 106; Indels 33; Gaps 10;

QY 29 KYLDQVTPKTFDTETIKELSRKQYSQASVLVGDIKVLLMSDQDKYFEATQVYEWCG 88  
Db 1369 KYISDAL--RLREBEKVVEEKQEHVKVKGWVSTLARTQK--ATSETKST 1423  
QY 89 VVTQLLSAYILLFDEVNEKK---ASAKDILIRLDDGVNKLNEAKSILGSSQSFN--- 142  
Db 1424 DIEKAILEQQVLSEELTTKEQVSEAIKTSQIFLAKHG-HKLESEKQKI--SEQNALN 1480  
QY 143 -----NASGKLALDSOLTNDPSEKSSYFOSQVDRIRKE-----AYAGAAAGIVAGP 189  
Db 1481 KAYHDLCDGSANQJQQQLAQHQTEQKT--LQKQNTCHQOLEDLCSWVGQAERALAGH 1538  
QY 190 FGLIISYSIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVKQAKDIDAKLKLKLA-TEIA 248  
Db 1539 QGRITQODLSA---LQKNQSDLKLDQDDIQNRATSPATVVKVDIEGFMENQTKLSPRELT 1595  
QY 249 AIGE 252  
Db 1596 ALRE 1599

RESULT 34  
US-10-369-493-19762  
; Sequence 19762, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 19762  
LENGTH: 862  
TYPE: PRT  
ORGANISM: NO. US20030233675A1toc punctiforme  
US-10-369-493-19762

Query Match 6.8%; Score 103; DB 14; Length 862;  
Best Local Similarity 23.7%; Pred. No. 7.8;  
Matches 79; Conservative 41; Mismatches 111; Indels 102; Gaps 17;

QY 38 KTFDETIKELSRFKQYSQASVLVGDIKVLLMSDQDKYFEATQVYEWCGVVTOLLISAY 97  
Db 61 KLDRTEQFFQOPKVSQGTSSYILGRSLDTLLDRADYVRKEFQDEYI---SIEHLLAY 117  
QY 98 I-----LLFDEYNEKKASAKQDILIRLDDGVNKLNEAKSILGSSQ-SFNNASGKLL 149  
Db 118 AKDRFGKALFQEG-----LDSE--KLKDI IKQVRGSKVTDQNPGEKYE 161  
QY 150 ALDSQLTNDPSEKSSYFQ-----SQVDRIR-----KEAYA-G 180  
Db 162 ALB-KYGRDLTEAARGQLDPVIGRDEIRRTVQILSRRTKNNPVLIGEPGVGKTAIAEG 220  
QY 181 AAAGIVAG--PFGI-----IISYSIAAGVIEGKLIPELNDRLKAVQNFT---SLSVTVK 230  
Db 221 LAQRITAGDVPOSLKDKRKLISLDMGALITACAKPRGEFEELKAVLKEVTESGGNVLIFID 280  
QY 231 QANKDIDAALK-----LATEIAAIGEIKTETTT-----RF---YV 264  
Db 281 EIHVVAGATQGMADAGNLLKPLMARGELRCIGATTLDEYRKHIEKDAALERRFOQVYV 340  
QY 265 DYDDL--MLSLKGAAKKMTNCNEYQORHGKK 295  
Db 341 DQPSVEDSISILRLRER-----YENHHGVK 366

RESULT 35  
US-10-080-608A-66  
; Sequence 66, Application US/10080608A  
; Publication No. US20030198956A1  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/10/080,608A  
; CURRENT FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 66  
; LENGTH: 961  
; TYPE: PRT  
; ORGANISM: Bos taurus  
US-10-080-608A-66

Query Match 6.8%; Score 103; DB 14; Length 961;  
Best Local Similarity 22.0%; Pred. No. 9.1;  
Matches 70; Conservative 53; Mismatches 129; Indels 66; Gaps 14;

QY 11 EVVKSIAETADGALDFYNNKL-----DQVTPKTFDE-----TIKELSRFK 51  
Db 639 EEVKKTLQEHDSIVTHYKNMIREQDLQLEELKQKISTLKQCNEQLQTAVTQVQSQIQOHK 698  
QY 52 QEYSQASVLVGDIKVLL--MDSQK--YFEATQVYEWCGVVTOLLISAYILLFDEVNEKK 108  
Db 699 DQYNL-----LKVQLGKOSQHQGPYTDGAQ-----MNGVQPEISR---LREEIEELK 743  
QY 109 AS-----AKDKILIRLDDGVNKK--LNEAQKSLGSSQSFNNAKGLALDSQLND 158  
Db 744 SNRELLQSLAEKSLIENLKSSQLSPGTNEQSSATAGDSEQIAELKQELATLKSQ-NS 802

|    |     |                                     |                              |     |
|----|-----|-------------------------------------|------------------------------|-----|
| Qy | 159 | FSEKSYFQSOVDRTKEAYACAAAGIYAGPGLII   | SYSTAAGVIEOKL-----IPELN      | 212 |
|    |     | ::: ::: ::: ::: ::: ::: ::: ::: ::: |                              |     |
| Db | 803 | QSVEITKIQTEQKQELQTEAFKAPVGESETVIAT  | KTD--VEGRSALLOETKELK         | 860 |
|    |     | ::: ::: ::: ::: ::: ::: ::: ::: ::: |                              |     |
| Qy | 213 | DRKAVONFTSLSVTVQKANKIDAAKLKLA       | TEIAAIGEIKTETETRTFYVDYDDMLMS | 272 |
|    |     | ::: ::: ::: ::: ::: ::: ::: ::: ::: |                              |     |
| Db | 861 | NETKALSEERTAIKQLDSSNTI-----AILQ     | NEKNKLEVDITDSKKEQDQLLV--     | 910 |
|    |     | ::: ::: ::: ::: ::: ::: ::: ::: ::: |                              |     |
| Qy | 273 | LKGAAGKMNINTCNEYQQ                  | 290                          |     |
|    |     | ::: ::: ::: ::: ::: ::: ::: ::: ::: |                              |     |
| Db | 911 | LLADODOKIFSLKNKLKE                  | 928                          |     |
|    |     | ::: ::: ::: ::: ::: ::: ::: ::: ::: |                              |     |

```

RESULT 36
US-10-370-685-155
; Sequence 155, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF.P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 155
; LENGTH: 961
; TYPE: prt
; ORGANISM: Bos taurus
US-10-370-685-155

```

|                       |        |                  |                 |                     |
|-----------------------|--------|------------------|-----------------|---------------------|
| Query Match           | 6.8%;  | Score 103;       | DB 14;          | Length 961;         |
| Best Local Similarity | 22.0%; | Pred. No. 9.1;   |                 |                     |
| Matches               | 70;    | Conservative 53; | Mismatches 129; | Indels 66; Gaps 14; |

  

|    |     |   |
|----|-----|---|
| Qy | 11  | EVKSAJETADGALDFYKYL-----DQVIPWKTFDE-----TIKELSRFK 51              |
| Db | 639 | EEVKKTLQHDHSIVTHYKIMIREQDLQLEELKQI:STLKQNEQLQTAVTQVQSQIQHQH 698   |
| Qy | 52  | QEYSQEASVLGDIKVLL--MDSODK--YFEATQTQVYEWCGVVTQLLSAYILLDFEYNEKK 108 |
| Db | 699 | DQYNL-----LKVQLGKDSQHGPYTDGAQ---MNGVQPEEISR--LREEIBELK 743        |
| Qy | 109 | AS-----AQKDILIRILDGYNK--LNEAQKLLGSSQSFNNASGKLALQSOLTND 158        |
| Db | 744 | SNRELLQSQAELKDSL:ENLKSSQLSPGTPNEQSSATAGDSQIAELKQELATLKSQ:-NS 802  |
| Qy | 159 | FSEKSSYFQSQVDRIKREAYAGAAGIVAGPFGILISYSIAAGVIEGL-----IPELN 212     |
| Db | 803 | QSVETIKLQTEKQBQLLQKTEAFKSAVPVGESETVIATKTTD--VEGRLSALLOETKELK 860  |
| Qy | 213 | DRLKAVNQFFTSLSVTVKQANKDIDAAKLKLATBATAIGEIKTETETTRFFYVDVDDLMS 272  |
| Db | 861 | NEIKALSERTAIKEQLDSSNSTI-----AILQNEKNKLEVDITDSKEQDDLLV- 910        |
| Qy | 273 | LLGAAKAKKMINTCNEYQQ 290   |
| Db | 911 | LLADODOKIFSLKNKLKE 928  |

RESULT 37  
US-10-408-765A-1635  
; Sequence 1635, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fany, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Taylor, Steven W.

```

; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1635
; LENGTH: 1583
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-408-765A-1635

Query Match          6.8%; Score 103; DB 16; Length 1583;
Best Local Similarity 19.9%; Pred. No. 18;
Matches 73; Conservative 55; Mismatches 89; Indels 150; Gaps 16

Qy      7 EQTVVVKSAIETADGALDFVNKYLDQVIPWKT--FDTIKELSRFKQEYQSEASVLVGD 64
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
531 EQTIT-----YNSELEQKVNELTGLEETLKE---KDNDQKLEKLMVQ 571

Qy      65 IKVLLMDSQDKYFEATQTVVWCWGVTQLLSAYI-LLFDEYNEKKAQAQKILIRILD-- 121
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:|
572 MKVL--SEDK-----EVLSEA VKSLYEENN--KLSEKKQLSRDLEV 610

Qy      122 -----DGVNKL--NEAQKSLGSSQGFFNNAAGSKLLALD 152
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:|
611 LSQKEDVILKEHITOLEKKQLMVBEQDNLNKLLENEQVKLFVTKLY----GFLKENG 666

Qy      153 SOLTNDPSEKSYPQSQVDRIKRAYAGAAAGIVAGPFGLIIYSIAAGVIEGKL----- 207
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:|
667 SEVSSEDSEKD-----VVNVLAQVGESLAKINEEK 697

Qy      208 -----TPELNDRIKAVQNFTPSLVTVKQANKIDAAKLKLTATIAIG----- 251
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:|
698 NLAQRDEKVLLEKETKCLQEEVSWQCEELKSLRDYEQEKKVLRKLEEIQSEKALQ 757

Qy      252 ----EIKTETTTRFYDYDDLMLSLKGAAKMINTCNEYQORHGK-----KT 296
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:|
758 SDLLEMKANAEKTR--LENQNLITQV-----EVVSQTCKSEIHNEKEKCPTKEHENLK 810

Qy      297 LLEVPI 303
Db      ||||:|
811 LLEOKEL 817
Db      ||||:|

```

RESULT 38  
US-10-282-122A-55283  
; Sequence 55283, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23





FEATURE:  
NAME/KEY: PEPTIDE  
LOCATION: (1869)..(1872)  
OTHER INFORMATION: anchor motif, sorting signal, cell wall targeting  
OTHER INFORMATION: region  
US-10-766-993-3

Query Match 6.8%; Score 102; DB 17; Length 1903;  
Best Local Similarity 19.1%; Pred. No. 28;  
Matches 63; Conservative 42; Mismatches 109; Indels 116; Gaps 12;

QY 6 ABQTVVVKSAIETADGALD-----FYNKYLDQVWPWT-----FDETTEL 47  
DB 1225 SEEKQAFQKALDNKAKALDNSETTEAEYKSNDELQAKADLDQQTDDKSKLDADA 1284  
QY 48 SRFK--QEYSQBASVLVGDIKVLLMDSQDKYFEATQTVYEWGVVTVLLSAYILLFBN 105  
DB 1285 NNAKGTDKYNASD-----DTKSKFDALKA-----BEVK 1315  
QY 106 EKASAKDILIRIILDDGVNKLNEAKSLGSS-----QSFNNASGKLLALDSQL 155  
DB 1316 NNSNATQKEV-----DDATNNLQKQNNLNGQTTDKSKLDADAIDANNAG 1361  
QY 156 TNDPSEKSYFQSVQVDRIRKEAYAGAAAGIVAGPFLIISYSIAGVIEGKLIPELNDRL 215  
DB 1362 TDYKNASDDTKSKFDALKA-----BEVKNS 1390  
QY 216 KAVQNPFTSLVTVVQAKNDIDAALKLATEATAIGETIKETETTRF---YVDYDLM 272  
DB 1391 NATQKEVDATNNLQAQNDLQ-----TTDKSKLDEAITDANNTKLTDKYNASDDTKS 1446  
QY 273 LKGAKKMINTCNEVQQRHGKKTLEVPD 302  
DB 1447 KFDEALKKAENKND-----SNATQKEVDD 1471

RESULT 44  
US-10-032-585-7611  
Sequence 7611, Application US/10032585  
Publication No. US20030180953A1  
GENERAL INFORMATION:  
APPLICANT: Terry, Roemer D.  
APPLICANT: BO, Jiang  
APPLICANT: Charles, Boone  
APPLICANT: Howard, Bussey  
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
FILE REFERENCE: 10162-005-999  
CURRENT APPLICATION NUMBER: US/10/032,585  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 8000  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 7611  
LENGTH: 1948  
TYPE: PRT  
ORGANISM: Candida albicans  
US-10-032-585-7611

Query Match 6.8%; Score 102; DB 14; Length 1948;  
Best Local Similarity 21.9%; Pred. No. 29;  
Matches 60; Conservative 41; Mismatches 97; Indels 76; Gaps 11;

QY 25 DFYNKYLDQVWPWTDETTELKLSRFQEQYSQBASVLVGDIKVLLMDSQDKYFEATQ--- 81  
DB 1434 DYIQLYDIT---KTLKSTREELNGSKTEILR-----LKALLRESBELYQVQKQENY 1482  
QY 82 --TVYEWGVVTVL-----LSAYILLFDEYNEKKAQKQILIRIL 120  
DB 1483 KTSVHDYEQDLQALQVKVKTLLSRNKDINSLRISYIKRSDEYTKKLELAESAIAISKRIE 1542  
QY 121 DDGVNKLNEAKSLGSSQSFNNASGKLLALDSQLNDFSEKSYFQSVQVDRIRKEAYAG 180  
DB 1543 EQATKEMKESRQALLVREB-----LRTTQILKDFRIKVENLEATIE----- 1586

QY 181 AAGAGIVAGPFLIISYSIAGVIEGKLIPE-LNDRKAVON---FFTSLSVTVQKANKDID 237  
DB 1587 -----KNHOLDANKKEIKQIQKLNHYLKNFENKELNEKKEIKNLRDLD 1633  
QY 238 AAKLKLATEATAIGETIKETETTRFVVDYDLM 271  
DB 1634 -----FKTDI-----ETKLIKENKKQLQDYEDVILL 1658

RESULT 45  
US-10-171-311-164  
Sequence 164, Application US/10171311  
Publication No. US20030087270A1  
GENERAL INFORMATION:  
APPLICANT: Schlegel, Robert  
APPLICANT: Chen, Yan  
APPLICANT: Zhao, Xumei  
APPLICANT: Monahan, John  
APPLICANT: Kamatkar, Shubhangi  
APPLICANT: Glatt, Karen  
APPLICANT: Gannavarapu, Manjula  
APPLICANT: Hoerish, Sebastian  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
TITLE OF INVENTION: OF CERVICAL CANCER  
FILE REFERENCE: MRI-035  
CURRENT APPLICATION NUMBER: US/10/171,311  
CURRENT FILING DATE: 2002-06-12  
PRIOR APPLICATION NUMBER: US 60/298,159  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/298,155  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/335,936  
PRIOR FILING DATE: 2001-11-14  
NUMBER OF SEQ ID NOS: 238  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 164  
LENGTH: 1938  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-171-311-164

Query Match 6.7%; Score 101.5; DB 14; Length 1938;  
Best Local Similarity 21.8%; Pred. No. 32;  
Matches 69; Conservative 48; Mismatches 105; Indels 95; Gaps 14;

QY 7 EQTVVVKSAI--ETAD-----GALDFYNKYLDQVWPWTDETTELKLSRFK 51  
DB 1173 EQEVTVLKALDEETRSHEAQVQEMRQKHAQAVEELTEQLEQ-----FKRAKANLDKNK 1226  
QY 52 QEYSQBASVLVGDIKVLLMDSQDKYFEATQTVYEWGVVTVLLSAYILLFDEYNEKKA 111  
DB 1227 QTEKENADLAGELRVL-----GQAKQEV-----EHKKKLEA 1259  
QY 112 QKDILIRILDG-----VNKLNEAKSLGSSQSFNNASGKLLALDSQLNDFSE 161  
DB 1260 QVQELQSKCSQGERARAEKLVKLVQNEVESVTG---MLNEAEGKAI---KLAKDVAS 1312  
QY 162 KSSYFQSVQVDRIRKEAYAGAAAGIVAGPFLIISYSIAGVIE---GKLIPELNDRLKAV 218  
DB 1313 LSSQLQDTQELLQEBTQK-----LNVSTKLQLEBEERNSLOQDLDEEMAK 1359  
QY 219 QNF---FTSLSVTVQKANKDIDAALKLATEATAIGETIKETETTRFVVDYDLM 275  
DB 1360 QNLEHISTLNIQSDSKKLIQ-----DPASTVEALEEGK-----KRFQKEIENLTQVVEE 1410  
QY 276 GAA---KQINTCNEYQQ 290  
DB 1411 KAAAYDKLEKTKNRLQQ 1427

RESULT 46  
US-09-927-597-2

```
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FASTSEQ FOR Windows Version 4.0
; SEQ ID NO 162
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-162
```

  

|  |                       |                      |                |                  |
|--|-----------------------|----------------------|----------------|------------------|
|  | Query Match           | 6.7%; Score 101.5;   | DB 14;         | Length 1972;     |
|  | Best Local Similarity | 21.8%; Pred. No. 33; | Mismatches 69; | Conservative 48; |
|  |                       |                      |                | Gaps 14;         |

  

```
QY      7 EQTVFVVKSAI--ETAD-----GALDFYNKYLDQVIPWKTDETIKELSRFK 51
          |||||:::||:||||::||::||::||::||::||::||::||::||:
Db     1173 EGEVTVLKKALDSETRSHQAQMQRKHAAVEELTEOLEQ-----PKRANLDKNK 1226

QY      52 QEYSQEASVLVDGIKVLLMDSODQKYFEATQTYYEWCGVVTTQLLSAVILLFDENYEKKASA 111
           |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~||:|:|
Db     1227 QTLEKENADLAGLRVL-----GQAKOEVS-----EHKKKLEA 1259

QY      112 OKXDLIRILDGG-----VNKNIEAQKSILGSSQSFNNASGKLALLDSLTDNFSE 161
         |::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~||:|:|
Db     1260 QVOEQSKSCSGGERAPAEINLVKHKLQNVESTGT---MLNEABEKAI----KLAKDVAS 1312

QY      162 KGSYFSQSDVRIRKEAYAGAAAIVAGPFGLIISYSIAAGVIH---GKIPELNDRLKAV 218
        ::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~||:|:|
Db     1313 LSSOLQDTQELLQEETRQK-----LNVSTKLRLQEEERNSLQQDQDDEWEAK 1359

QY      219 QNF---FTSLSVTVMQANKIDDAUKLATEIAAIGEIKTETPTTPFYDYDDIMLSLLK 275
        :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~||:|:|
Db     1360 QNLERHIISTINTIQLSDSKKKLQ---DFASTVEALEEGK-----KRFOKEIENTOQYEE 1410

QY      276 GAA--KKMINTCNEYOO 290
        ::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~||:|:|
Db     1411 KAAAYDKLETKNRLOO 1427
```

  

RESULT 48  
 US-10-341-434-103  
 ; Sequence 103, Application US/10341434  
 ; Publication No. U520030215835A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Origene Technologies  
 ; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes  
 ; FILE REFERENCE: 9U 204 205 R1  
 ; CURRENT APPLICATION NUMBER: US/10/341,434  
 ; CURRENT FILING DATE: 2003-07-18  
 ; PRIOR APPLICATION NUMBER: US 60/348,164  
 ; PRIOR FILING DATE: 2002-01-15  
 ; PRIOR APPLICATION NUMBER: US 60/348,119  
 ; PRIOR FILING DATE: 2002-01-15  
 ; NUMBER OF SEQ ID NOS: 238  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 103  
 ; LENGTH: 1972  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens
 US-10-341-434-103

  

|  |                       |                      |                |                  |
|--|-----------------------|----------------------|----------------|------------------|
|  | Query Match           | 6.7%; Score 101.5;   | DB 14;         | Length 1972;     |
|  | Best Local Similarity | 21.8%; Pred. No. 33; | Mismatches 69; | Conservative 48; |
|  |                       |                      |                | Gaps 14;         |

  

```
QY      7 EQTVFVVKSAI--ETAD-----GALDFYNKYLDQVIPWKTDETIKELSRFK 51
          |||||:::||:||||::||::||::||::||::||::||:
Db     1173 EGEVTVLKKALDSETRSHQAQMQRKHAAVEELTEOLEQ-----PKRANLDKNK 1226

QY      52 QEYSQEASVLVDGIKVLLMDSODQKYFEATQTYYEWCGVVTTQLLSAVILLFDENYEKKASA 111
           |:|:|:|:|:|:|:|:~::~||:|:|
Db     1227 QTLEKENADLAGLRVL-----GQAKOEVS-----EHKKKLEA 1259
```

```
QY 112 QKDLIRILDG-----VNKLEAQKSLGSSQSFNNASGKLLALDSQLTNDPSE 161
Db 1260 QVOELQSCSDGERARAEINLKVHKLQNEVESVTG---MLNEAEGKAI---KLAKDVAS 1312
QY 162 KSSYFQSDVRIRKEAYAGAAAGIVAGPFGLLISYSIAAGVIE---GKLIPELNDRLKAV 218
Db 1313 LSSQLOQTQELLQEBTRQK-----LNVSTKRLQLEERNLSLQDLDEMEAK 1359
QY 219 QNF---FTSLSVTVKQANKDIDAAKLKLATEIAAIGETKTETTRFYVDYDDMLSLK 275
Db 1360 QNLERHISTLNIQLSDSKKLQ-----DFASTVEALEEGK-----KRFQKEIENLTQOYEE 1410
QY 276 GAA---KWMINTCNEYQQ 290
Db 1411 KAAAYDKLEKTKNRLQQ 1427

RESULT 49
US-09-927-597-4
; Sequence 4, Application US/09927597
; Publication NO. US20030032018A1
; GENERAL INFORMATION:
; APPLICANT: Malik, Fady
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; APPLICANT: Craven, Andrew
; APPLICANT: Sakowicz, Roman
; APPLICANT: Hartman, James
; TITLE OF INVENTION: Human smooth muscle myosin heavy chain
; FILE REFERENCE: CYTOPO18
; CURRENT APPLICATION NUMBER: US/09/927,597
; CURRENT FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Human
US-09-927-597-4

Query Match 6.7%; Score 101.5; DB 10; Length 1979;
Best Local Similarity 21.8%; Pred. No. 33;
Matches 69; Conservative 48; Mismatches 105; Indels 95; Gaps 14;

QY 7 EQTVVVVKSAT--ETAD-----GALDFYKYLDOVIPWKTFTDETIKELSRPK 51
Db 1180 EQBTVLTKALDETSHERAQVQEMRQKHAQVVEELTEQLQ-----FKRAKANLDRNK 1233
QY 52 QEYSQASVLVGDIKVLLMDSODKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKASA 111
Db 1234 QTLKENADLAGELRVL-----GQAKQEV-----EHHKKKLEA 1266
QY 112 QKDLIRILDG-----VNKLEAQKSLGSSQSFNNASGKLLALDSQLTNDPSE 161
Db 1267 QVOELQSCSDGERARAEINLKVHKLQNEVESVTG---MLNEAEGKAI---KLAKDVAS 1319
QY 162 KSSYFQSDVRIRKEAYAGAAAGIVAGPFGLLISYSIAAGVIE---GKLIPELNDRLKAV 218
Db 1320 LSSQLOQTQELLQEBTRQK-----LNVSTKRLQLEERNLSLQDLDEMEAK 1366
QY 219 QNF---FTSLSVTVKQANKDIDAAKLKLATEIAAIGETKTETTRFYVDYDDMLSLK 275
Db 1367 QNLERHISTLNIQLSDSKKLQ-----DFASTVEALEEGK-----KRFQKEIENLTQOYEE 1417
QY 276 GAA---KWMINTCNEYQQ 290
Db 1418 KAAAYDKLEKTKNRLQQ 1434
```

```
GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6362
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6362

Query Match 6.7%; Score 101; DB 14; Length 1475;
Best Local Similarity 17.9%; Pred. No. 24;
Matches 53; Conservative 52; Mismatches 117; Indels 74; Gaps 8;

QY 7 EQTVVVVKSATIE-----TADGALDFYKYLDOVIPWKTFTDETIKELSRPKQEYS 55
Db 231 EESIKLLSQIELLRQSHSQSLSEIQEARIFEKMLTQQV-----DSAMKAKSDREAAK 285
QY 56 QEASVLVGDIKVLLMDSODKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKASAKQDI 115
Db 286 AREQVLEKQVQELRLKLEEDPEKNQLVHNLALNAQI-----EELTKAL 331
QY 116 LIRILDDGVNKLNEAQAQSLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSDVRIRK 175
Db 332 KVDSMQOGATASEDRIRELIGGHQEAIKQLENTKMMNESLQDLVEKEARFSEMERIRT 391
QY 176 EAYAGAAGIVAGPFGLLISY-----STAAGVIBGKL-----IPELNDRL 215
Db 392 ESQTTSES-----LKYEHVLVKMMIEMDEKLEAEVLAKSQOANLEIQEPHDKI 441
QY 216 KAVQNFFTSLSVTVKQANKDIDAAKLKLATEIAA-----IGEIKTETTRTF 262
Db 442 KQLE-----LEVQLSSENKEKIQAEMLMVQEKASENIKNAEKKVNGLEAEVKEKLR 492

Search completed: January 28, 2005, 19:28:34
Job time : 148 secs
```





Wed Jan 5 14:16:58 2005

544 KSTKSEYVWMLDEANYDITCLRNKNDKLEAVNKYRECESKET 587

Db 120  
 QY 180  
 Db 180  
 QY 240  
 Db 240  
 QY 300  
 Db 300  
 QY 305  
 Db 305

RESULT 2

US-10-437-963-197045

Sequence 197045, Application US/10437963

Publication No. US20040123343A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A.

APPLICANT: Barbazuk, Brad

APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

Plants and Uses Thereof for Plant Improvement

TITLE OF INVENTION: 38-21(53221)B

FILE REFERENCE: 38-21(53221)B

CURRENT FILING DATE: 2003-05-14

CURRENT FILING DATE: 2003-05-14

NUMBER OF SEQ ID NOS: 204966

SEQ ID NO 197045

LENGTH: 889

TYPE: PRT

ORGANISM: Oryza sativa

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT4530\_92840C.1.pap

US-10-437-963-197045

Query Match 8.6%; Score 130.5; DB 16; Length 889;  
 Best Local Similarity 22.1%; Pred. No. 0.068;  
 Matches 76; Conservative 66; Mismatches 123; Indels 79; Gaps 17;  
 QY 8 QTVVVKSAIETADGALD-----LYNKYLDQVVPKTFDETIK-----ELSRKQY 54  
 Db 268 EKVEILSSVRVLKGLDSTABESKRETEELV--KNLESEVSVLKGLSEARIEER 325  
 QY 55 SQEASVLVGIKVLMDSDQKFEATQTVYEH--CGVVTQLLSAYILLPDEY-NEKKAS 110  
 Db 326 LATEKLIBELKSEVADAKKABSEARQLFEWKHAGLEMELEA-VTLSDKFKGESLAS 384  
 QY 111 AQDKILRIILDDGVKKLNEAQSLLTSQSFFNNAAGKLLALD---SQTNDPSEKSYF- 166  
 Db 385 T-----TEELGKIQSALQDRESEIEVLGKTKTALIEVARLLADVNESNQFD 432  
 QY 167 -----SQVDRIKAEYAGAAAGVAGPFGLLIISYSIAAGV-IEG-----KLPIE 210  
 Db 433 ASQOEVEGLQTTIDVLNKLKLEAAEAASEA-----LNNEKANKVIEGLTEENVKLI 486  
 QY 211 LN-----NRLKTVNFFTSATVKQANKDIDAALKLATE-----IAAIGIKET 257  
 Db 487 LNETRDREKRAVEDDITAAALS-----BESDKAEAHERVLSKEDDHEHALAQIGDKMAL 543  
 QY 258 ETTR----FYVDYDDLMLSLKGAACKVINTCNEYQOR-HGKKT 296

Db

RESULT 3

US-10-369-493-10918

Sequence 10918, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

PLANTS WITH IMPROVED PROPERTIES

TITLE OF INVENTION: 38-10(52052)B

FILE REFERENCE: 38-10(52052)B

CURRENT FILING DATE: 2003-02-28

CURRENT FILING DATE: 2003-02-28

PRIOR FILING DATE: 2002-02-21

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 10918

LENGTH: 891

TYPE: PRT

ORGANISM: Ferroplasma acidarmanus

US-10-369-493-10918

Query Match 8.3%; Score 126; DB 14; Length 891;

Best Local Similarity 22.8%; Pred. No. 0.16;

Matches 71; Conservative 52; Mismatches 119; Indels 70; Gaps 13;

QY 4 IPAEQTVEV-----VKSAIETADGALDLYNKYLDQVVPKTFDETIK 46

Db 505 IFGSEIETASYNIEKSLTEKKELENTIENASKGHELFNVLSAEN---ENLEKAIKE 561

QY 47 LSRFKQYVQSEASVLVGDIKVLLMD--SQDKYFEAT-QTVYEWCGVVTQLLSAYILLFDE 103

Db 562 LQYENYIRYSIIISG-----MDPGAKEIEAAGESFTTPKRNINELLS-QIGFVPE 614

QY 104 YNEKASAKQDILIRILDDGVKKLNEAQSLLTSQSFFNNAAGKLLALDLSQTLNDFSEKS 163

Db 615 YNEYQNTGKISSEINRLKTEVERSRMKSLSKLEIKDEIENKKSIEGLRIEMENKQSAWH 674

QY 164 SYFQSOVDRIKAEYAGAAAGVAGPFGLLIISYSIAAGVIEG--KLPIELNRLKTVQNF 221

Db 675 QY-----DGIDEQ-----AGNIESRYKSAQENIKRKTIVDS 706

QY 222 FTS-LSATVKQA-NKDIDAALKLATE-IAAIGIKETETETFRFVYDDVDDLMLSLKGA 278

Db 707 YTERIETENAKNLQDAEKYKKTREATISTIGKI-----REAFDINGIOSIIRK 759

QY 279 KKMINTCNEYQQ 290

Db 760 ASMTNLTRKYLQ 771

RESULT 4

US-09-884-696-5

Sequence 5, Application US/09884696

Publication No. US20030035809A1

GENERAL INFORMATION:

APPLICANT: GEORGE, LISA W

APPLICANT: ANGELOS, JOHN A

APPLICANT: HESS, JOHN F

TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES

TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA

TITLE OF INVENTION: BOVIS INFECTIONS

FILE REFERENCE: 481-06

FILE REFERENCE: 481-06

CURRENT FILING DATE: 2001-06-19

CURRENT FILING DATE: 2001-06-19

NUMBER OF SEQ ID NOS: 41

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 5

```
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-884-696-5

Query Match      7.5%; Score 116; DB 10; Length 1023;
Best Local Similarity 20.8%; Pred. No. 1.3;
Matches 69; Conservative 59; Mismatches 111; Indels 98; Gaps 11;

Qy 2 TSFAETQVTVVK-----SAIETADGALDLYNKYLDQVWPWKTPTDETIKLSRPKQE 53
Db 155 TALSSMKIDELIKKQSGGVSSSELAKASIELINQLVDTA-----ASLNNVNSFSQQ 207
Qy 54 YSQEASVL-----VGDIKVLLMDSQKYFEATQTVYEWGCVVTVLLSAYILL-PDEY 104
Db 208 LNKLGSLVSLNTHLNGVGN-KLQNLPLNDINAGLDTV---SGILSAISAFILSNADAD 263
Qy 105 NEKASAKQDILIRILD---GVKKNEAQKSLTSSQSFNNASGKLLALDSQLTNDFSE 161
Db 264 TGTKAAGVELTKVLGVNGKISQYIIAQAAGLST-----302
Qy 162 KSSYFOSQVDRIRKEAYAGAAAGIVAGPFGLIIS-----YSIAAGVIEGKLIPELNNRLKT 217
Db 303 -----AAAAGLIASVWTLAISPLSFLSIADKFKRANKIEVSQRFKK 344
Qy 218 VQNFPTSLSATVKQANKDIDAALKLATEIAAIGEIKTETETTRFYVDYDDMLSLKGA 277
Db 345 LGYDGSLLAAFHKETGAIDASLSTRIVLASVSSGISAAATTSLV---GAPVSAVLGA 400
Qy 278 -----AKKMINTCNEYQQRHGK 294
Db 401 VTGIISGILEASKQAMFEHVASKWADVIAEWKKGK 437

RESULT 5
US-10-369-493-5220
; Sequence 5220, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5220
; LENGTH: 2823
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5220

Query Match      7.5%; Score 113.5; DB 14; Length 2823;
Best Local Similarity 20.8%; Pred. No. 8.5;
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;

Qy 25 DLNVKYLQVTPW-KTFDETIKELSRFKQBSQASVLVG-----DIKVLMLDS 72
Db 1881 DLKNR-IDVLEQWMDYRETIYDVSKDADARMSLVGKRNRYKEVSNIEKLRVEA 1939
Qy 73 QDK-----YFEATQTV-----YEWGCVVTVLLSAYILLFDEY 104
Db 1940 EDQIAYSRNSIEKARSELMMFEDKEKINWTLAELPDLVEQCONITLL---YSQLIDEY 1996
Qy 105 NEK-----KASAKQDILIRILD-----DGVKKLENAQKS 133
Db 1997 DEEYVQTAGRHAELKLEVQAQK-IVDRFVDTRTETENPLKASHAYENIVEALKNATEAVDS 2055
Qy 134 LLTSSQSFNNASGKLLALDSQLTNDFSEKSSYFOSQVDRIRKEAYAGAAAGIVAGPFGLI 193
Db 2056 AAEEASE-----AVSKMLGSESGSDANEES--LRSQLEKLNKSSLSN-----2097
Qy 194 ISYSIAAGVIE--GKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAALKLATEIAAIG 251
Db 2098 VDNSNAVKIVEELKKEKKDITDLGHLNELKTSI-----VKRLG 2136
Qy 252 EIKTETETTRFYVDYDDMLSLKGAAGKM-INTCNEYQQRHGKKT 296
Db 2137 VIKNEASS---WDDKDRMHSILKNGAKTAHERSANNVKESEGIKT 2179

RESULT 6
US-10-369-493-5221
; Sequence 5221, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5221
; LENGTH: 2823
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5221

Query Match      7.5%; Score 113.5; DB 14; Length 2823;
Best Local Similarity 20.8%; Pred. No. 8.5;
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;

Qy 25 DLNVKYLQVTPW-KTFDETIKELSRFKQBSQASVLVG-----DIKVLMLDS 72
Db 1881 DLKNR-IDVLEQWMDYRETIYDVSKDADARMSLVGKRNRYKEVSNIEKLRVEA 1939
Qy 73 QDK-----YFEATQTV-----YEWGCVVTVLLSAYILLFDEY 104
Db 1940 EDQIAYSRNSIEKARSELMMFEDKEKINWTLAELPDLVEQCONITLL---YSQLIDEY 1996
Qy 105 NEK-----KASAKQDILIRILD-----DGVKKLENAQKS 133
Db 1997 DEEYVQTAGRHAELKLEVQAQK-IVDRFVDTRTETENPLKASHAYENIVEALKNATEAVDS 2055
Qy 134 LLTSSQSFNNASGKLLALDSQLTNDFSEKSSYFOSQVDRIRKEAYAGAAAGIVAGPFGLI 193
Db 2056 AAEEASE-----AVSKMLGSESGSDANEES--LRSQLEKLNKSSLSN-----2097
Qy 194 ISYSIAAGVIE--GKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAALKLATEIAAIG 251
Db 2098 VDNSNAVKIVEELKKEKKDITDLGHLNELKTSI-----VKRLG 2136
Qy 252 EIKTETETTRFYVDYDDMLSLKGAAGKM-INTCNEYQQRHGKKT 296
Db 2137 VIKNEASS---WDDKDRMHSILKNGAKTAHERSANNVKESEGIKT 2179

RESULT 7
US-10-369-493-1061
; Sequence 1061, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
```

APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 1061  
LENGTH: 1005  
TYPE: PRT  
ORGANISM: Methanococcus jannaschii  
US-10-369-493-1061

Query Match 7.5%; Score 113; DB 14; Length 1005;  
Best Local Similarity 19.7%; Pred. No. 2.3;  
Matches 60; Conservative 55; Mismatches 91; Indels 98; Gaps 12;

QY 7 EQTEVVKSAIETADGALDYNKYLDQVVPKWTDE-----TIKELSRP-----KQYSQ 56  
DB 605 DEILEDIKSOLNFKP--NFYNOYLSAVSYLNSVDEEGRNRRIKEIENIVSGWNKEKRE 661  
QY 57 EASVL-----VGDIKVLMDSDQKYPEATQTVYEWCGVVTQLLSAVILLFDEYNKAS 110  
DB 662 ELNKLREDEREINRLKOKNELKKEKE-----LIEENRSLNFKDYKEYLGL 710  
QY 111 AQKDLIRILDGKVKLNEAOKSLTSSQFNNSGKLLALD----- 152  
DB 711 TEKLELNKIDGLEEI-----YNIENRSLNFKDYKEYLGL 757  
QY 153 -----SQTNDPSEKSSYFQSDVDRIRKAYAGAGIVAGFGLIISYIAAGVIEGKI 208  
DB 758 ILEVNEKINDIEBERISYNOKLDE-----INNEE-----EHKI 792  
QY 209 PEL--NNR--LKTQNFFTSLSATVKQANKDIDAKLKLATEIAAIGBIKTETTFYV 264  
DB 793 KELYENKQELDNVREKQTEITGTGLEYLKKOVESLAKL-----KEMSNLEKEKEKLTQFV 848  
QY 265 DYDD 268  
DB 849 EYLD 852

RESULT 8  
US-10-424-599-212070  
Sequence 212070, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 212070  
LENGTH: 564  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_33525C.1.pap  
US-10-424-599-212070

Query Match 7.4%; Score 112.5; DB 15; Length 564;  
Best Local Similarity 20.7%; Pred. No. 1.2;  
Matches 62; Conservative 57; Mismatches 110; Indels 71; Gaps 12;

QY 11 EVVKSAIETADGALDYNKYLDQVVPKWTDETIKELSRFKQYSQASVLVGD---IKV 67  
DB 61 EVEREKLRAEANLE-----KQAMDWMLAQEELKRLGEDAARHAEBSESSETLEDFFRRVKK 114  
QY 68 LMDSDQKYPEATQTVYEWCGVVTQLLSAVILLFDEYNKASQAQ-----KDLIRILD 121  
DB 115 LLNDRSELVSSQALASSRSMKEBERLLELQELSELGEQASVMSYMNENKDAQIEV-- 172  
QY 122 DGVKVLNEAOKSLTSSQFNNSGKLLALDSQTLNDFSEKSSYFQSDVDRIRKAYAGA 181  
DB 173 -----ESERTKLRAESRRELERDLKMEKELISELEER-----LKKE----- 210  
QY 182 AAGIVAGFPGLIISYIAAGVIEGKIPELANRLKTVQNFFTSLSATVKQANKDIDAKL 241  
DB 211 -----RTSLQAVKREVALLOEELK-KTAE--PRETSAVLQVKESELVDAKL 254  
QY 242 KLATEIAAIGBIKTETTFYVDYDDLMLSLLKGAACKMINTCNE--YOOR---HGKKT 296  
DB 255 E-----IQLKSEKASLOGLIEKDELELS-----SARKMLGDVYQEIYDLKMLMHSKET 303

RESULT 9  
US-10-032-585-7611  
Sequence 7611, Application US/10032585  
Publication No. US20030180953A1  
GENERAL INFORMATION:  
APPLICANT: Terry, Roemer D.  
APPLICANT: Bo, Jiang  
APPLICANT: Charles, Boone  
APPLICANT: Howard, Bussey  
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
FILE REFERENCE: 10182-005-999  
CURRENT APPLICATION NUMBER: US/10/032,585  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 8000  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 7611  
LENGTH: 1948  
TYPE: PRT  
ORGANISM: Candida albicans  
US-10-032-585-7611

Query Match 7.4%; Score 112.5; DB 14; Length 1948;  
Best Local Similarity 22.2%; Pred. No. 6.2;  
Matches 58; Conservative 40; Mismatches 90; Indels 73; Gaps 10;

QY 38 KTFDETIKELSRFKQYSQASVLVGDIKVILLMDSDQKYFEATQ-----TVYEWCGVVTQ 92  
DB 1444 KTLKSTREELNGSKTEILR-----LKALLRESEDELYQVKQENYKTSVHDYEQDLAQ 1495  
QY 93 L-----LSAVILLFDEYNKASQAQKDILI--RILDDGVKVLNEAOKS 133  
DB 1496 LKVKHETLLSRNKDINESLIYKKRDEYVKKLELAESAIAISKRHEEQATKEMKESRSQ 1555  
QY 134 LLTSSQFNNSGKLLALDSQTLNDFSEKSSYFQSDVDRIRKAYAGAGIVAGFGLI 193  
DB 1556 LLAVREE-----LRTTQILIKDPRIKVENLEATIEE----- 1586  
QY 194 ISYIAAGVIEGKIPE--INNRLKTVQ--FFTSLSATVKQANKDIDAKLKLATEIAAI 250  
DB 1587 KNHOLDANKKEIKQIQDKLNLKPNFENKELKEEIKNLNRDLD-----FKTDI--- 1638  
QY 251 GEIKTETTFYVDYDDLM 271  
DB 1639 -ETKLEKNKKQLDYEDVLL 1658

RESULT 10  
US-10-282-122A-70920  
Sequence 70920, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-07-27

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-09-06

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-09-09

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-09-06

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-09

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 70920

; LENGTH: 1189

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

; US-10-282-122A-70920

Query Match 7.4%; Score 112; DB 15; Length 1189;

Best Local Similarity 18.3%; Pred. No. 3.5;

Matches 61; Conservative 60; Mismatches 130; Indels 82; Gaps 9;

Qy 7 EQTEVVVK-----SATEADGALDLYNKYLDQVWPWKTDETIKELSRKQYQSASV 60

Db 163 EESAGVLKYYKKAESQKLDHTEDNLNRVEDILYDLGRVPLKEEAIAKEYQLSKE 222

Qy 61 LVGDIKVLMDSDQKYFEATQTVYEWG-----GVVTTLSAYILLFDEYNE 106

Db 223 MEQSDVITVSDIDHYTDDNQRLDERLNLKSOAEKEGQQAQINQLQRY----- 273

Qy 107 KKAQAQKILIRILDGQVKLNEAQAQSLTSSQFNNAAGKLLALDSQTLND-----FS 160

Db 274 -KGRQON-----DYDIEKLN---YELVKATENEYQSLGNVLBERKKNQSETNARYE 323

Qy 161 EKSSYFQSDVRIRKEAYAGAAIGVAPGLIISYSIAAGVIEGKLIPELNNRL-----K 216

Db 324 EELNLESQSDISKNEKAQNE-----KLLADLKNKQKQLNK 359

Qy 217 TVQNFFTSLSATVQKANDDAKILATETAATGEIKTETTRFFVVDYDMLSLKLG 276

Db 360 EVOELESLLYVSDQHDKEELIKNSYITLMSQSDVNNDIRFLEHTINENAKKSRIDS 419

Qy 277 AAKMINTCNEYQ-----RHGKTKLFEV 300

Db 420 RLVEAFNLQDKIQNTQTNKAYQSSKSMQKV 452

# RESULT 11

US-10-282-122A-56861

; Sequence 56861, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-07-27

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-09-06

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-09-09

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-09-06

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-09

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 56861

; LENGTH: 1158

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

; US-10-282-122A-56861

Query Match 7.4%; Score 111.5; DB 15; Length 1158;

Best Local Similarity 21.0%; Pred. No. 3.7;

Matches 66; Conservative 58; Mismatches 121; Indels 69; Gaps 15;

Qy 4 IFAEQTEVVVVKSAIETADGALDLYNKYLDQVWPWKTDETIKELSRKQYQSASV 63

Db 183 VIAQONVN-----QITDTOL-LYQNQLEB-----TIQPPKPEIPMNYTTVLSTASILDE 230

Qy 64 DIKVLMDSDQKYFEATQTVYEWGCVVTTLSAYILLFDEYNKKAQAKDILIRILDG 123

Db 231 DNK-LFTNEQEAQVSKVQT-----LLENSRSLESTSKTETDKSV-----DEY 274

Qy 124 VKKLINE-AQKSLTSSQFNNAAGKLLALDSQTLNDSEKSSYFQSQVDRIRKEAYAGAA 182

Db 275 SKANEKIKKSIQAQNEQPERQK-----QELTNQWESDTTVYKQFQDLN----- 319

Qy 183 AGIVAGPFGLIISYS-----IAAGVIEGKLIPEL-NNRLKTVQNFFTSLSATVQANK 234

Db 320 -GNVINQFSFYTPSEGGSGIYADFLSESKLFQETQGNRIGELQKEIAELHMQVEQ--- 375

Qy 235 DIDAAKLKLA-----TEIAAIGEIKTETTRFFVVDYDMLSLKGAAGKMWINTCNEYQ 290

Db 376 -LTALKQIAATYNDLEATPEIATDTQIKQ-----AIIQITNEKENIPNLDKNYQE 427  
Qy 291 RHGKKTLEFEVDVA 304  
Db 428 RL-EESLSEISFVS 440

RESULT 12

US-10-028-248A-107  
; Sequence 107, Application US/10028248A  
; Publication No. US20030235882A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Vernet, Corine  
; APPLICANT: Casman, Stacie  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Gangolli, Esha  
; APPLICANT: Miller, Charles  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Li, Li  
; APPLICANT: Taupier Jr, Raymond J  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Zethusen, Bryan  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Coleman, Steven  
; APPLICANT: Tchernev, Velizar  
; APPLICANT: Si, Jingsheng  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Stone, David  
; APPLICANT: Sciore, Paul  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Rothenberg, Mark  
; TITLE OF INVENTION: NO. US20030235882A1el Nucleic Acids and Polypeptides and Methods  
; TITLE OF INVENTION: Thereof  
; FILE REFERENCE: 21402-222  
; CURRENT APPLICATION NUMBER: US/10/028,248A  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/256619  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/262959  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: 60/272408  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/285189  
; PRIOR FILING DATE: 2001-04-20  
; PRIOR APPLICATION NUMBER: 60/308039  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: 60/311266  
; PRIOR FILING DATE: 2001-08-09  
; NUMBER OF SEQ ID NOS: 211  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 107  
; LENGTH: 1999  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-10-028-248A-107

Query Match 7.4%; Score 111.5; DB 14; Length 1999;  
Best Local Similarity 19.6%; Pred. No. 7.8;  
Matches 60; Conservative 49; Mismatches 114; Indels 83; Gaps 11;  
Qy 7 EQTVVEVKSATF-----TADGALDLYNKYLDQVIP--WKTFTDET---IKELSRFKQEVQSQE 57  
Db 116S EQEVNLLKTLBEEAKTHEAQIQEMRQKHSQAVEIAEQLEQTKRKVKANLEKAKOTLENE 1224  
Qy 58 ASVLVGDIDKVLMSDQKYFEATQVYEWCGVVTQQLSAYILLDFEYNEKKSAAQKOILI 117  
Db 122S RGLANEVKKVLQGGED-----SEHKRKKVQAQLQELQ 1257

Qy 118 RILDDGVKKLNEAKSKLLTSSQSFNNASCKLLALD---SOLTNDSEKSSYFQSQVDRLR 174  
Db 1258 VKFNEGERVTLADKVKTLQVELDNVTGLLSQSDSKSKLTKDFSALQSQDQTQELQ 1317  
Qy 175 KEAYAGAAAGIVAGPEGLIISYSIAAGVIEGKLIIPELNNRLKTQV---NFTTSLSATVK 230  
Db 1318 EEN-----RQKL--SLSTKLAQVEDEKNSFREQLSEEBE 1349  
Qy 231 QANKDIDAARKLATEIAAIGIKTETETTFYVDYDDL-MLSLKGAAKKMMINTCNEYQ 289  
Db 1350 EAKHNLEK---QIATLHAQVADMKKQME-----DSVGCLETAEEVKKRKLQKQDLEGLS 1398  
Qy 290 QRHGKK 295  
Db 1399 QRHEEK 1404

RESULT 13

US-10-107-782-107  
; Sequence 107, Application US/10107782  
; Publication No. US20040018970A1  
; GENERAL INFORMATION:  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Casman, Stacie  
; APPLICANT: Coleman, Steve  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Gangolli, Esha  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Li, Li  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Miller, Charles  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Rothenberg, Mark  
; APPLICANT: Sciore, Paul  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Si, Jingsheng  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Stone, David  
; APPLICANT: Taupier, Raymond, jr.,  
; APPLICANT: Tchernev, Velizar,  
; APPLICANT: Vernet, Corine,  
; APPLICANT: Zethusen, Brian  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF  
; FILE REFERENCE: 21402-222CIP  
; CURRENT APPLICATION NUMBER: US/10/107,782  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 10/028,248  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/256,619  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/262,959  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: 60/272,408  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/285,189  
; PRIOR FILING DATE: 2001-04-20  
; PRIOR APPLICATION NUMBER: 60/308,039  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: 60/311,266  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 60/279,344  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 215  
; SOFTWARE: Curaseqdist version 0.1  
; SEQ ID NO 107  
; LENGTH: 1999  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-107-782-107



Db 875 ALASQL-EEIQLKLVGESQVENKEEL-----ICAKIMNKEMVD 913  
Qy 210 ELNNRLKTVQNFFTSLSATVKQAKOIDAAKKLATEIA-----AIGEIKT--- 255  
Db 914 ELNAKLGDAEGMBELKKSLEVSSEAKVORREBELIAQVSKHRDQEQQLTLDELKSAQH 973  
Qy 256 ETETTR 261  
Db 974 STETSR 979

RESULT 16  
US-10-369-493-5109  
; Sequence 5109, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 5109  
; LENGTH: 1938  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-5109

Query Match 7.3%; Score 110; DB 14; Length 1938;  
Best Local Similarity 18.9%; Pred. No. 9.9;  
Matches 55; Conservative 60; Mismatches 114; Indels 62; Gaps 9;  
Qy 5 FAEQTVVVK-----SAIETADGALDLYNKYLQVDPKTFDETIKELSRFK 51  
Db 1315 FSSQLVEAKAAEDLHERQEFHACKNLEHLDQCHELLEEQINGK--DDIQRLSRIN 1372  
Qy 52 QEYSQ-----EASVLVG--DIKVLMDSQKYFEATQTVYEWCGVVTTOLLSAYILLFDEY 104  
Db 1373 SEISQWKARYEGEGLVGSSELEELKQKQNRVMDLQEALSAQNKVILEKAKGKLLAET 1432  
Qy 105 NEKKAQAQDILI-----RILDDGVKLNKAEQKSLTSSQSFNNAAGKLLALD 152  
Db 1433 EDARSDVDRHLTVIASLEKKQKQAFDKIVDDWKRVDDIQKEIDATTRDSRNTSTEVFKLR 1492  
Qy 153 SOLTNDPSEKSSVFQSDVRIRKEAYAGAAAGIVAGPFGLLIISYSIAAGVIE-----GKL 207  
Db 1493 SSMND-----LSQIETLARE-----NKIFSQEI-RDINEQITQGGRT 1529  
Qy 208 IPELNNRLKTVQNFFTSLSATVKQAKOIDAAKKLATEIAAIGIKTETE 258  
Db 1530 YQEVHKSVRLEQEKDELQHALDEAALAEABESKVLRLQIEVQVQIRSEIE 1580

RESULT 17  
US-10-369-493-5110  
; Sequence 5110, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 5110  
; LENGTH: 1938  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-5110

FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 5110  
LENGTH: 1938  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-369-493-5110

Query Match 7.3%; Score 110; DB 14; Length 1938;  
Best Local Similarity 18.9%; Pred. No. 9.9;  
Matches 55; Conservative 60; Mismatches 114; Indels 62; Gaps 9;  
Qy 5 FAEQTVVVK-----SAIETADGALDLYNKYLQVDPKTFDETIKELSRFK 51  
Db 1315 FSSQLVEAKAAEDLHERQEFHACKNLEHLDQCHELLEEQINGK--DDIQRLSRIN 1372  
Qy 52 QEYSQ-----EASVLVG--DIKVLMDSQKYFEATQTVYEWCGVVTTOLLSAYILLFDEY 104  
Db 1373 SEISQWKARYEGEGLVGSSELEELKQKQNRVMDLQEALSAQNKVILEKAKGKLLAET 1432  
Qy 105 NEKKAQAQDILI-----RILDDGVKLNKAEQKSLTSSQSFNNAAGKLLALD 152  
Db 1433 EDARSDVDRHLTVIASLEKKQKQAFDKIVDDWKRVDDIQKEIDATTRDSRNTSTEVFKLR 1492  
Qy 153 SOLTNDPSEKSSVFQSDVRIRKEAYAGAAAGIVAGPFGLLIISYSIAAGVIE-----GKL 207  
Db 1493 SSMND-----LSQIETLARE-----NKIFSQEI-RDINEQITQGGRT 1529  
Qy 208 IPELNNRLKTVQNFFTSLSATVKQAKOIDAAKKLATEIAAIGIKTETE 258  
Db 1530 YQEVHKSVRLEQEKDELQHALDEAALAEABESKVLRLQIEVQVQIRSEIE 1580

RESULT 18  
US-10-282-122A-53523  
; Sequence 53523, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931



**Prior Filing Date:** 2000-12-22  
Prior Application Number: 60/267,636  
Prior Filing Date: 2001-02-09  
Prior Application Number: 60/269,308  
Remaining Prior Application Data removed - See File Wrapper or PALM.  
**Number Of Seq ID NOS:** 78614  
Software: PatentIn version 3.1

**Seq ID NO 53523**

Type: PRT  
Organism: Clostridium difficile

US-10-282-122A-53523

Query Match      7.2%; Score 109.5; DB 15; Length 742;  
Best Local Similarity    21.4%; Pred. No. 3;  
Matches         66; Conservative 49; Mismatches 131; Indels 63; Gaps 10;

QY 18 ETADGALDLYNKVLDQVPWKTFDTIKELSRFKOYEYSQAASVLGVGIKLVLMDSQKYF 77  
DB 236 KTVDLASDATSKLSDTVDIKNLSLTFTTKTT-----LNDDTKLLSSDKPKFLFDNDNLD 288  
QY 78 EATQTVEWCVVTOLLSAYILLPEYNKKASAOKILIRILD-----DGVKKLNEAQ 131  
DB 289 ELSPKSLDNLMVDL-----SSASSUTNLIDAVNSGSDEPVKLLIDLNL 333  
QY 132 KSILTSSQSFFNNAGSKGLALDSLQN-----DFSEKSYFSQSVDRIRKEAYA 179  
DB 334 SEKSNLQLSNTLDVFTLKQLTSNNRLDVINDLEDSSNKIDSSTISLANDINKVIS 393  
QY 180 GAAAGIVAGPFGIIISYSTIAAGVIGKIPIELNNR-L-KVNFTTSLSATVKQAKDI-- 236  
DB 394 GOOPSISA--LNVLSLSNGINGRIINLINLPFNPSKISKIPINNIFAN--SIKVANDITV 448  
QY 237 --DAKULK-----LATETAATGEIKETETTTRFRFYVDVDDLMLSLKGAAKMINTCNEY 288  
DB 449 LKAERAPKPVBELLTTSUKSGNAQESILIR-----BRLPLAKGMDDLDITLUKI 501  
QY 289 QQRHGKKTL 297  
DB 502 SNGEDMKKL 510

**Result 19**  
US-10-437-963-191043  
Sequence 191043, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
Applicant: La Rosa, Thomas J.  
Applicant: Kovalic, David K.  
Applicant: Zhou, Yihua  
Applicant: Cao, Yongwei  
Applicant: Wu, Wei  
Applicant: Boukharov, Andrey A.  
Applicant: Barbazuk, Brad  
Applicant: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(5322)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 191043  
LENGTH: 815  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_8739C.1.pap  
US-10-437-963-191043

Query Match                  7.2%; Score 108.5; DB 16; Length 815;  
Best Local Similarity       20.5%; Pred. No. 4.1;  
Matches                 71; Conservative 69; Mismatches 121; Indels 85; Gaps 16;



APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 497  
LENGTH: 344  
TYPE: PRT  
ORGANISM: Xenorhabdus nematophilus  
US-10-369-493-497

Query Match 7.0%; Score 106.5; DB 14; Length 344;  
Best Local Similarity 21.2%; Pred. No. 1.9;  
Matches 56; Conservative 52; Mismatches 103; Indels 53; Gaps 13;  
QY 15 SAETADGALDLYNKYLQVDPKWT--FDTEIKELSRFKQESQASVLVDIKVLMDS 72  
DB 71 TVLHVGDSDSIRLHPQDISDLPKLGIDIVLDTGKYGERIDGEGAHLASGAKKVLFAHP 130  
QY 73 QDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASAKDI--LIRILDGQVKKLE 129  
DB 131 GGNLDLAT-VY--GVNHLLSA-----EDCIVNSACTNCIPIIKVLDLDD----- 174  
QY 130 AQKSLTSSQFNNAKGLALDQSLTNDPSEKSSYPQSQVDRIKRAYAGAAAGIVAGP 189  
DB 175 -----AFNIESGTVTTHASM-NDQPVIDAY-HSDLRTR-----AASQSIIP-- 215  
QY 190 FGLIISYIAAGVIEGKLIPELNNRLKTQVNTFTSLATVKQANKOIDAAKCLKLATEIAA 249  
DB 216 -----VPTKLAAGIT--RIPKFNRPFAISVRVPTINVTAILDLSVTSA-----AVNVSE 264  
QY 250 IGBI--KTETETFRFYVDYDML 271  
DB 265 VNQLLQKSAVSPRGIVDFDLPL 288

RESULT 24  
US-10-425-114-51665  
; Sequence 51665, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 51665  
; LENGTH: 647  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700221631\_FLI.pep  
US-10-425-114-51665

Query Match 7.0%; Score 106.5; DB 15; Length 647;  
Best Local Similarity 22.8%; Pred. No. 4.4;  
Matches 79; Conservative 52; Mismatches 121; Indels 95; Gaps 15;  
QY 9 TVEVVKSAIETADGALDLYNKYLQVDPKWTFTETIKELSRFKQESQASVLVDIKVL 68  
DB 79 TVQLEKARYELANAIDAKNEALSQV-----DDAV-----RASEAKTQVEHLIAVTHL 128

QY 69 --LMDS-----QDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASAKDI--LIRIL 120  
DB 129 KGLVDSKVDGNRKKTATRIQNLEE-----ENFALKLELQKKAAEKAIRLERMI 178  
QY 121 DD-----GVK-----KLENAQKSLTSSQFNNAKGLALD 152  
DB 179 DELQSDVDDARNFGSKQLADEWQKQALLEVLEEDQSNILKGSLSNAMESLSDTS 238  
QY 153 SOLTNDPSE-----KSSYFQSQVDRIKRAY-----AGAAAGIVAGPGLIISYSTAAG 201  
DB 239 SLLRDRESEVAALRDKVRFLDELTLKNDIVVSGQRADAAEKAADLWTEVEGLRLKLH 298  
QY 202 VIEGKLIPELNNRLKTQVNTFT--SLSATVKQANKOIDAAKCLKLATE--IAAIGEIK 254  
DB 299 TVSEKIEALNSDKSLSLETETLNEQKNQKLADLEANKD--ELEKTKAMGLASALQEMS 357  
QY 255 TET-----ETTRFYVDYDMLSLKGLAAKMKMINTCNEYQ 289  
DB 358 AESREAREKYLKQDIERAQAQVEELSLSLK-----NTKENYE 396

RESULT 25  
US-10-369-493-9374  
; Sequence 9374, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 9374  
; LENGTH: 857  
; TYPE: PRT  
; ORGANISM: Xylella fastidiosa  
US-10-369-493-9374

Query Match 7.0%; Score 106.5; DB 14; Length 857;  
Best Local Similarity 24.3%; Pred. No. 6.4;  
Matches 69; Conservative 45; Mismatches 103; Indels 67; Gaps 14;  
QY 7 EQTVVVKSAIETADGALDLYNK-----YLDQVIFPKTFTETIKELSRFKQESQASV 60  
DB 574 EEAIVVSDAVRRSRTGLSDPNRPSGFLPGTGVGKT--ELCKALAEFLD--SODAMV 630  
QY 61 LVGDIKVLWDSODKYFEAT-----QTVYEWCGVVTQLLSA--VILLFDEYNEKAS 110  
DB 631 -----RIDMSEFWEKHSVARLIGAPGVYEGGVTLELRRRPSYLSILLDEV-EKAHS 684  
QY 111 AQXDILIRILDGQVKKLENAQKSLTSSQFNNAKGLALDQSLTNDPSEKSSYFQSQV 170  
DB 685 DVFNILLQVLDDG--RLTDQGRTV---DPRNT---VIVMTSLGSHQIQLSGDDSD-- 733  
QY 171 DRIRKAYAGAAAGIVAGPGLIISYIAAGVTEGKLIPELNNRLKTQVNTFTSLATVK 230  
DB 734 ---PEVYTMKAAMV-----GVVQAHPRPEFINRLDDIVVFHPLDKAQIK 775  
QY 231 QANKOIDAAKCLKLATEIAAIGEIKTETETTRFYVDYDMLSL 274  
DB 776 Q-----IARQLRGLERKLAESLUK-----LDLDDRALELL 806

RESULT 26  
US-10-393-602-148  
; Sequence 148, Application US/10393602

Publication No. US20030170714A1  
GENERAL INFORMATION:  
APPLICANT: Gregory Dolganov  
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory Polypeptides  
NUMBER OF SEQUENCES: 151  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/393,602  
FILING DATE: 19-Mar-2003  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,126  
FILING DATE: 26-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 148:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1312 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54  
SEQUENCE DESCRIPTION: SEQ ID NO: 148:  
US-10-393-602-148  
Query Match 7.0%; Score 106.5; DB 14; Length 1312;  
Best Local Similarity 18.9%; Pred. No. 11;  
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;  
QY 10 VEVVKSATETADGALDLYNKYLDQVVPWKTPDETIK-----ELSRFKQEY 54  
DB 502 MEVISIQNEKAD--LDRTLRKLDQEMEQLNHHHTTTRTQMEMLTQDKADKDEQIRKIKSRH 559  
QY 55 SEQASVLVGDGKVLNMDSDQKYEATQTYVWCGVVTQLLSAYILLFDYNEKKASAKQD 114  
DB 560 SDELTSLG-----YFPNKKQLEDWLHKSCK-----EINQTR----- 591  
QY 115 ILIRILDGDKVKKLNEAKQSILTSQSFNNASGKLALDLSQLTNDFSEK-----SSYFQS 168  
DB 592 -----DRLAKLN---KELASSEQNKHNNELKREBQLSS-YEDKLFVCGSQDFES 640  
QY 169 QVDRIRKE-----AVAGAAAGIVAGFF----- 190  
DB 641 DLDRLKEEIEKSKORAMLAGATA--VYSQFTQLTDFENOSCPVCQVQRFQTEAEQLQEV 698  
QY 191 -----GLI-ISYSIAAGVIEGKLIPELNNRKT 217  
DB 699 SDIQSKLRAPDKLKSTESLKKKKRDEMLGLVPMRQSIID--LKEKEIFELRNKLQN 756  
QY 218 VQNFFTSLSATVQANKNDAAKLKLAETAAIGRIKTETTT-----RPYVDY 266  
DB 757 V-----NRDIQRLKNDIBEQETLLGTIMPBEESAKVCLDVTIMERFQML 802  
QY 267 DDLMLSLLKGAAK-----KMINTCNEYQQRHGKK 295  
Db 803 KDVERKTAQQAALQGIDLDLDRTVQOVNQEKQKH 838  
RESULT 27  
US-09-885-535-4  
; Sequence 4, Application US/09885535  
; Patent No. US20020104105A1  
; GENERAL INFORMATION:  
; APPLICANT: Myriad Genetics, Inc.  
; APPLICANT: Heichman, Karen  
; APPLICANT: Bartel, Paul L.  
; TITLE OF INVENTION: Protein-Protein Interactions  
; FILE REFERENCE: 2318-266-II  
; CURRENT APPLICATION NUMBER: US/09/885,535  
; CURRENT FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 60/213,245  
; PRIOR FILING DATE: 2000-06-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 4  
; LENGTH: 2835  
; TYPE: PPT  
; ORGANISM: Homo sapiens  
; US-09-885-535-4  
Query Match 7.0%; Score 106; DB 9; Length 2835;  
Best Local Similarity 19.4%; Pred. No. 36;  
Matches 74; Conservative 76; Mismatches 122; Indels 110; Gaps 17;  
QY 7 EQTVVVK-----SAIETAD-----GALDLYNKYLDQV-IPWKT 39  
DB 807 KSTVEIKREGEKIAATPAADKVKILQSLDSRWEALLNKATRNQLEGISVVAQQ 866  
QY 40 FDTIKELSRFK-----QYSQSEASVL---VGDIKVLLMDSQKYEATQTVIEW 86  
DB 867 FHETLEPLNEWLTTEKRLVNCPEPIGTQASKEEQIAQHKVLQEDILLRKQNVQDALLNG 926  
QY 87 CGVVTQLLSAYILLFDYNEKKASAKQD-----LIRILDGV---KKLNEAQKSLITS 137  
DB 927 LELLKQTTGDEVLI IQDKLEAIKARYKDITKLTSTDAKTLEQALQALRLHSTHELCT- 985  
QY 138 SQSFNNASGKLALDLSQLTNDFSEKSSYFQSQVDRIRKEAYAGAAAGIVAGPFLIISYS 197  
DB 986 --WLDKVEVELLSYETQVLK--GEEASQAMRPEKELKKEAKNKA-----LIDSLN 1032  
QY 198 -IAAGVIE-----GKLIPELNNRKTQVNFSTLSATVQKANKDIDA----- 238  
DB 1033 EVSSALLELVPMRAREGKEMVAEDNERYLV-----SDTITQKVEIDAILRSQOF 1085  
QY 239 -----AKKLATE-----IAAGIKETETETTFVVD-----YDDLMLSLKGA 278  
DB 1086 DQADAELSWITTEKKIMSLGDIRLEODTSALQVQKTFMTEILRHKDIIDDLVK-SG 1144  
QY 279 KMINTCNEYQQRHGKKTLEFV 300  
DB 1145 HKINTACSEEEKQSMKKLDKV 1166  
RESULT 28  
US-10-408-765A-666  
; Sequence 666, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Watnook, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME

[illegible]

```

QY 173 IRKENYAGAAAGIVAGPFGGLIISV-IAAGVIE-----GKLIPELNNRLKTVQNF 221
      :|||      |      :|||      :|||      :|||      :|||      :|||      :|||
Db 912 LKKEAKNNKA-----LLDSLNEVSSALLLVLPWRAREGLEKVAEDNERYELV--- 959

QY 222 FTSLSATVKQANKDIDA-----AKLKATE-----IAAIGEIKTETETTRFYV 264
      :|||      :|||      :|||      :|||      :|||      :|||      :|||      :|||
Db 960 ----SDTITQVEISDAILLRSQQFDQAADAELSWITETEKKLMISGLDIRLEQDOTSAQL 1015

QY 265 D-----YDDLMLSLKGAAKWINTCNEYQOORHGKTKLFPV 300
      :|||      :|||      :|||      :|||      :|||      :|||      :|||      :|||
Db 1016 QVQKTFMTWELRHKDIIDDLVK-SGHKIMTACSEEEKQSMKKKLDKV 1061

RESULT 31
US-10-425-115-226623
; Sequence 226623, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 226623
; LENGTH: 584
; TYPE: PRI
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_138267C.1.pep
US-10-425-115-226623

Query Match 6.9%; Score 104.5; DB 17; Length 584;
Best Local Similarity 20.7%; Pred. No. 5.6;
Matches 73; Conservative 67; Mismatches 109; Indels 103; Gaps 18;

QY 7 EQTVVVKSAIEATDAGLDLYNKYLDQVTPWKTPTETIKELSRKQEV-----SOEAS 59
Db 187 EDELQEARQOIE-----ALEYKNHCO-----CEKLESILKQVSEKEDLVAILASKNEVE 238

QY 60 VLVGDIKVLMDSDQKYFEATQVY-----EMCGVVVTQLLSAYILLDFEYNEKKSAAQKD 114
Db 239 DLKGD-----MVSAKHFEA-QLVHRDHEIEKCKQAEQVS-----EKYPHEKSTLSE 286

QY 115 I-----LIRILDGVKKLN-----BAQ-KSLTSSQSFNNASGKLLA-----LDSOL 155
      :|||      :|||      :|||      :|||      :|||      :|||      :|||      :|||
Db 287 IERLQEVVRGFEENLTAVAGEKQLEAQVKELEQTSNDLDDSSAIIKLOIHKDLQGLR 346

QY 156 TNDPSEKSSYFO--SQVDRIRKAYAGAAAGIVAGPFGGLIISYSAAGVIEGKLIPELNN 213
      :|||      :|||      :|||      :|||      :|||      :|||      :|||      :|||
Db 347 ENDSNDKSVLEERAMELEQVRRQLEDSRAEASMK---LQTTIKNLNVLEEK--AELEN 400

QY 214 RLKTVQNFPTSLSATVKQANKDIDAKKLI-----ATEI-AAIGEIKT 255
      :|||      :|||      :|||      :|||      :|||      :|||      :|||      :|||
Db 401 RMKDAEQATSDLNSLVGSLDGKLKATEAQLEQLHAEKABAILSEKQISELNQAIRUKT 460

QY 256 ETEITRFVVDYDDLMLSLKGA-------KKMINTCNEYQOORH 292
      :|||      :|||      :|||      :|||      :|||      :|||      :|||      :|||
Db 461 RLE-----LLSSEKAAVDTKASALLTDVAVRDEKLTDRHLQQLH 501

RESULT 32
US-10-425-114-64494
; Sequence 64494, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.

```



Matches 60; Conservative 63; Mismatches 120; Indels 76; Gaps 14;

7 EQTEVVVKSATETAD---GALD-----LYNKYLDQVVPKTFDTIKE-----LS 48

192 ETTVNDLRKSVESWTLSGLGAVENSITQATLSNIQQTKELSNTRNISEMLDILSGMS 251

49 RFQOYSQASVLVDGIKVLMDSDQKYFEATQTVYEWCVVTOILSAYILLFDEYNEKK 108

252 NIQEISETKAFGGKLHSEVIEDK-MDAT-----NQLITS---KPFDEFSELL 297

109 ASAQKDLIRLDDGVKVLNEAQKSLLT-----SSQSFNNASGKLIALDSQLT 156

298 KKSNTALVEVMKVTVEEFORQNSLNKLQENFDOLNKSVEKLTWQOENKAMISLT 357

157 NDFSEKSSYFOS---QVDRIRKEAYAGAAAGIVAGPPGLIISYSIAAGVIEGKLIPELNN 213

358 QOYKEMASNPSTSTTLISQVGGDTRTLVSEG---GKLKQLID-SLNQVIVEDQKFDVSN 413

214 RLKTVQNFPTS-----LSATVQANKDIDAACKLKLATEIAAIGIKTETETTRF 262

414 KLOETANISKEMKPFDESTKILNDWVRKORNFVGVQL-LIEKLDLKNIR----- 464

263 YVDYDMLMLSLKGAACKM 281

465 --DYSE---QFWKGTKEKM 478

RESULT 36

US-10-282-122A-51808

Sequence 51808, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Lianguo

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zvekind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITEA 034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 51808

LENGTH: 809

TYPE: PRT

ORGANISM: Clostridium acetobutylicum

US-10-282-122A-51808

Query Match 6.9%; Score 104; DB 15; Length 809;

Best Local Similarity 22.3%; Pred. No. 9.6;

Matches 67; Conservative 40; Mismatches 110; Indels 84; Gaps 14;

24 LDLYNKYLDQVVPKTFDTIKELSRFBQBYSEASVLVDGIKVLMDSDQKYFEATQTV 83

265 LVLYKRNQNVYDFSTYQNAINDLNNKLAFSQO-----KI-----SNKY-----T 305

84 YEWCGV-----VTOLLGAYILLFDEYNEKASQAQKDLIRLDDGVKVLNEAQKSLTSS 138

306 YEWFSIPVNVQVQDLKSQVHYTDE-----BIDSLLR--DGGLKIYTTMTSMESNV 355

139 QSP--NNASGKLIALDSQLTNDSEKSSYFOSQVDRIRKEAYAGAAAGIVAGPPGLIISY 196

356 QNILDNNSTLK-----SYSDAKNGIIQPEAAATLFDVHTGEIKAIIVGG----- 399

197 SIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVQANKDIDAACKLKLATEIAAIGIKTE 256

400 -----RGOQPPSSYNRADS--SNYLSRVGSSSIKPLTVYAPAITKLTATEDTIVNDSPLS 451

257 TETTRFY---VDY-----DDLMLS---LLKGAACKMIN-----TCNEYQQRH 292

452 SDVAEKYGSNGVVPYHPHNDGGYSGPVNLTALTKSINLVAIKLEDKLGISTGAAYAOKF 511

293 G 293

512 G 512

RESULT 37

US-10-369-493-3279

Sequence 3279, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 3279

LENGTH: 1965

TYPE: PRT

ORGANISM: Neurospora crassa

US-10-369-493-3279

Query Match 6.9%; Score 104; DB 14; Length 1965;

Best Local Similarity 21.2%; Pred. No. 32;

Matches 63; Conservative 45; Mismatches 119; Indels 70; Gaps 11;

6 AEQTVVVVKSATETADGALDLYNKYLDQVVPKTFD-----ETIKELSRFBQBYSQ 56

57 AKYREELAKQLEHAKEKT-----NDWLDNELTKVQAEAKFKKKGARIABLQRLNEDANS 112

57 EASVLVDGIKVL---LMDSDQKYFEATQTVYEWG-----GVVTOLLSA--YILLFDE 103

113 TIESLTRSEQVLRKLIQEAQDKAEETLTKVQOQEAARTEEGFKQELSAKLVKLDQ 172

104 YNEKFKASQKDLIR---ILDDGVKVLNEAQKSLTSSQSFNNASGKLIALDSQLTNDFS 160

173 QSETHRNRLKEVLELRLEQIKDDHANEIRIRRELEQKE-----DHAQT---E 217



```
QY 161 EKSSYFOSQVDRIRKE-----AYAGAAAGIVAGPGLIISYSIAAG 201
Db 218 QRAQELQNEVDRIKASSDLGRSPGTPQPKGGFATRAG---SPFGTPIUSIRGRAG 274
QY 202 VIEGKLIPELNNRLKTVQNFPTSLSATVVKQANKDIDAAKLKLATEIAAIGEIKETETE 258
Db 275 QRATDALEELYN---VKQLAGEKRCCKLQBELDDAVAMLEAKWPEIDELNAESE 327

RESULT 38
US-10-425-115-267777
; Sequence 267777, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 267777
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(568)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_175816C.1.pcp
US-10-425-115-267777

Query Match 6.8%; Score 103.5; DB 17; Length 568;
Best Local Similarity 21.4%; Pred. No. 6.5;
Matches 70; Conservative 51; Mismatches 115; Indels 91; Gaps 14;

QY 3 SIFAEQTVVVKSAIETADGALDLYNKYLQVIP--WKTPETIKLSRFKQYEQEASV 60
Db 9 SVFASGTAEFLSTAQA-----DVANKALQDNTRKLTXTHTKTVTSALVK-TDTHLNEALS 62
QY 61 LVGDIKVLMDSQDKYFEATQTVYEWGVVTVLLSAYILLDFEYNEKKASAKQDILIRIL 120
Db 63 EISSLESGLQDAEKRF-----VTMQELRDYISVMCDFLNDKA-----FLIEEL 105
QY 121 DDGVKKLINEAQKSLTSSQSFNNSGKLLALDLSQLTNDFSEKSSYFQSQVDRIKAYAG 180
Db 106 EENIQQLHEK-----RALAISERRAADLADESGVI-----EAAVS 140
QY 181 AAAGIVA-GPFGLIISYS-----IAAGVIRG--KLIPEL-----N 212
Db 141 AAVSILSKGSSSCLSNASNAQAAMAAAGSSNLQPELDFGRDINMQKMDLKREED 200
QY 213 NRLKTVQNFPTSLSATVVKQANKDIDAAKLKLATEIAAIGIKETETETETTFYVDDMLMS 272
Db 201 RRQRKTQSETKRLASAAK--NKDIKKEGELSTD-----ESDSESTAVVSRDE---- 247
QY 273 LLKGAACKMINTCNEYQQRHGKKTLPF 299
Db 248 FLKAAADVFDIAKEEYSLSLRIVKDKPE 274

RESULT 39
US-10-408-765A-1635
; Sequence 1635, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
```

```
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1635
; LENGTH: 1583
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-408-765A-1635

Query Match 6.8%; Score 103.5; DB 16; Length 1583;
Best Local Similarity 19.7%; Pred. No. 26;
Matches 69; Conservative 54; Mismatches 89; Indels 139; Gaps 15;

QY 7 EQTVVVKSAIETADGALDLYNKYLQVIPWKT--FDETIKLSRFKQYEQEASVILVGD 64
Db 531 EQTIQ-----YNSELRQKVNELTGGLEETLKE-----KDQNDQKLEKLMVQ 571
QY 65 IKVLLMDSQDKYFEATQTVYEWGVVTVLLSAYI--LLDFEYNEKKASAKQDILIRILD-- 121
Db 572 MKVL-----SEDK-----EVLSAEVKSLYEENN--KLSSSEKKQLSRDLVLF 610
QY 122 -----DGVKKL--NEAOKSLTSSQSFNNSGKLLALD 152
Db 611 LSQKEDVILKEHTQLEKKLQMLMVEQDNLKLENEQVKLFVKTLQY-----GFLKEMG 666
QY 153 SOLTNDSEKSSYFQSQVDRIKAYAGAAAGIVAGPGLIISYSIAAGVIEGKL----- 207
Db 667 SEVSEDSSEKD-----VNVLQAVGSLAKINEKC 697
QY 208 -----IPELNNRLKTVQNFPTSLSATVVKQANKDIDAAKLKLATEIAAIG----- 251
Db 698 NLAQFQDEKYLEKEIKCLQESVVCBELKSLRLDYEQEKVLLRKELEFIOSEKEALQ 757
QY 252 ----BIKTETETTFYVDDMLMSLLKGAACKMINTCNEYQQRHGKKTLPF 298
Db 758 SDLLEMMNANEKTR--LENQNLLIQV-----EEVSQTCSEIHNEKEKCF 801

RESULT 40
US-10-437-963-124349
; Sequence 124349, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 124349
; LENGTH: 1585
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1585)
; OTHER INFORMATION: unsure at all Xaa locations
```

```

; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 36
; LENGTH: 1959
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-028-248A-36

Query Match      6.8%; Score 103.5; DB 14; Length 1959;
Best Local Similarity 19.2%; Pred. No. 35;
Matches 68; Conservative 49; Mismatches 119; Indels 119; Gaps 13;

QY 7 EQTVVVKSAIE-----TADGALDLYNKYLQDVIPWKTFTDIKE-----LSRFKQEYS 55
Db 1166 EQVNVLLKTLLEAKTHEAQIQEMRQKHSQAV--BELAEQLEQTKRVKANLEKAKOTLE 1223
QY 56 QFASVLVGDIKVLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKOI 115
Db 1224 NERGELANEVKKVLLQGGD-----SEHKRKVEAQLQE 1256
QY 116 LIRILDGCVKLNKAQSKLLTSSQSFNNASGKLLALD---SOLTNDPSEKSSVFSQSDVR 172
Db 1257 LOVFNEGERVTELADKVTKLQVELDNVTGLLSQSDSKSKLTDFSALESQLODQTEL 1316
QY 173 IRKEAYAGAAAGIVAGPFLIISYIAAGVIEGKLIPELNNRLKTVQ---NFF----- 222
Db 1317 LOEEN-----RQKL--SLSTKLQVDEKNSFREOLEEE 1348
QY 223 -----TSLSATV-----KQNKDIDAUKLATEIAAIG 251
Db 1349 EAKHLEKQIATLHAQVADMKKMEDSVGCLTAEVVKRLQKQLEGLSQRHEEKVAAYD 1408
QY 252 EIKTETETFRFYVDYDDMLSLKGAAGKMNITCN-EYQORHGKTKLFEVDPVAS 305
Db 1409 --KLEKTKTRLQOELDLVLDL-----DHQASACNLEKKQKQFDOLLAEKTTISA 1457

RESULT 42
US-10-107-782-36
; Sequence 36, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steve,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Gangolli, Esha,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Li, Li,
; APPLICANT: Liu, Xiaohong,
; APPLICANT: Malvankar, Uriel,
; APPLICANT: Miller, Charles,
; APPLICANT: Millet, Isabelle,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Sciore, Paul,
; APPLICANT: Shenoy, Suresh,
; APPLICANT: Shinkets, Richard,
; APPLICANT: Si, Jingsheng,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Stone, David,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Tchernev, Velizar,
; APPLICANT: Vernet, Corine,
; APPLICANT: Zerhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222CIP
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 10/028,248

; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_27097C.1.pap
; US-10-437-963-124349

Query Match      6.8%; Score 103.5; DB 16; Length 1585;
Best Local Similarity 22.1%; Pred. No. 26;
Matches 61; Conservative 48; Mismatches 114; Indels 53; Gaps 10;

QY 2 TSIFAEQ--TVEVVKSAIETADGALDLYNKYLQDVIPWKTFTDIKEKLSRFKQEYSQBAS 59
Db 608 TSLLAEKQHGAAIMSLAQETKQNEELQKLDVN---RTNDLQDSLKRPEENVVTRDA 664
QY 60 VLVG-----DIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKO 113
Db 665 LYLAEQHEBDETQSLSKSQERNWELLQKVDEAKINKLLENAQRL-----EKHATAKE 719
QY 114 DILIRIL---DDGVKVLNKAQSKLLTSSQSFNNASGKLLALD---SOLTNDPSEKSSY-- 165
Db 720 SLLKTKQSHDSTTKALVAEASRNRELTKSFEDSKINLLEDSVNRLEERIAEKDSLLE 779
QY 166 FQSOVDRIKAYAGAAAGIVAGPFLIISYIAAGVIEGKLIPELNNRLKTVQNFSTSL 225
Db 780 IERENNAATKDEVINA-----QNKIM-ELVNESQLOQDIRKHL 816
QY 226 SATVQKQNKDIDAUKLATEIAAIGELKUTETTR 261
Db 817 EDNIKRLEED-----ATTREALLISEKQTHEATR 846

RESULT 41
US-10-028-248A-36
; Sequence 36, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shinkets, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malvankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glenda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
```



|                             |  |
|-----------------------------|--|
| Publication No.             | US20030235882A1  |
| GENERAL INFORMATION:        |  |
| APPLICANT:                  | Shimkets, Richard  |
| APPLICANT:                  | Patturajan, Meera  |
| APPLICANT:                  | Vernet, Corine   |
| APPLICANT:                  | Casman, Stacie   |
| APPLICANT:                  | Malyankar, Uriel   |
| APPLICANT:                  | Shenoy, Suresh   |
| APPLICANT:                  | Spytek, Kimberly   |
| APPLICANT:                  | Gangolli, Bsha   |
| APPLICANT:                  | Miller, Charles  |
| APPLICANT:                  | Boldog, Ferenc   |
| APPLICANT:                  | Li, Li   |
| APPLICANT:                  | Taupier Jr, Raymond J  |
| APPLICANT:                  | Kekuda, Ramesh   |
| APPLICANT:                  | Smithson, Glennda  |
| APPLICANT:                  | Zerhusen, Bryan  |
| APPLICANT:                  | Liu, Xiaohong  |
| APPLICANT:                  | Colman, Steven   |
| APPLICANT:                  | Tchernev, Velizar  |
| APPLICANT:                  | Si, Jingsheng  |
| APPLICANT:                  | Edinger, Shlomit   |
| APPLICANT:                  | Stone, David   |
| APPLICANT:                  | Sciore, Paul   |
| APPLICANT:                  | Millet, Isabelle   |
| APPLICANT:                  | Rohenberg, Mark  |
| TITLE OF INVENTION:         | No. US20030235882A1el Nucleic Acids and Polypeptides and               |
| FILE REFERENCE:             | 21402-222  |
| CURRENT APPLICATION NUMBER: | US/10/028, 248A  |
| CURRENT FILING DATE:        | 2001-12-19   |
| PRIOR APPLICATION NUMBER:   | 60/256619  |
| PRIOR FILING DATE:          | 2000-12-19   |
| PRIOR APPLICATION NUMBER:   | 60/262959  |
| PRIOR FILING DATE:          | 2001-01-19   |
| PRIOR APPLICATION NUMBER:   | 60/272408  |
| PRIOR FILING DATE:          | 2001-02-28   |
| PRIOR APPLICATION NUMBER:   | 60/285189  |
| PRIOR FILING DATE:          | 2001-04-20   |
| PRIOR APPLICATION NUMBER:   | 60/308039  |
| PRIOR FILING DATE:          | 2001-07-26   |
| PRIOR APPLICATION NUMBER:   | 60/311266  |
| PRIOR FILING DATE:          | 2001-08-09   |
| NUMBER OF SEQ ID NOS:       | 211  |
| SOFTWARE:                   | PatentIn Ver. 2.1  |
| SEQ ID NO 103               |  |
| TYPE:                       | PRN  |
| ORGANISM:                   | Homo sapiens   |
| US-10-028-248A-103          |  |
| Query Match                 | 6.8%; Score 102.5; DB 14; Length 1961;                                 |
| Best Local Similarity       | 19.5%; Pred No. 42;  |
| Matches                     | 60; Conservative 47; Mismatches 114; Indels 87; Gaps 11;               |
| QY                          | 7 EQTVVEVKSIAIE----TADGALDLYNKYLDQVIPWKTFTETKE-----LSRPKQEYS 55        |
| Db                          | 1166 EQEVNLLKTTLEEAKTHEAQIQENRQKHSQAV--BELAEQLEQTKRVKANLEKAKQOTLE 1223 |
| QY                          | 56 QEASVLVGDIKVLMDSDQKYFEATQTVYEWCGVQTLLSAVILLDFEYNEKKASQAQKDI 1115    |
| Db                          | 1224 NERGELANEVKKVLLQGRD-----SEHKRKKEVAQLQE 1256                       |
| QY                          | 116 LTRILDGKVKLNEAKSLITSSQSFNNASGKLALD---SOLTNDPSKSSYFQSQVDR 172       |
| Db                          | 1257 LQVKNFEGEVRVELADKVTKLQVELDNTVGLLSQSDSKSKLTQDFALESOLQTOEL 1316     |
| QY                          | 173 IRKEAVAGAAGVIGFGLIISYAAGVIEGKLIPELNNRLKTVQ---NFTLSLSAT 228         |
| Db                          | 1317 LOEEN-----ROKL--SUSTKUKQVDEKNSFREQLLEE 1348                       |
| QY                          | 229 VKQANKDIDAAKLKLAETAAIGETETTFYFVYDDDL-MLSLKGAAKKMWTCNE 287          |

Db 1349 EBEAKHNLEK--QIATLHAQVADMKKME-----DSVGCLETAEEVKRLKQDLEG 1397  
QY 288 YQQRHGKK 295  
Db 1398 LSQRHEEK 1405

## RESULT 47

US-10-107-782-103  
; Sequence 103, Application US/10107782  
; Publication No. US20040018970A1  
; GENERAL INFORMATION:  
; APPLICANT: Boldog, Perenc,  
; APPLICANT: Casman, Stacie  
; APPLICANT: Colman, Steve,  
; APPLICANT: Edinger, Shlomit,  
; APPLICANT: Gangolli, Baha,  
; APPLICANT: Kekuda, Ramesh,  
; APPLICANT: Li, Li,  
; APPLICANT: Liu, Xiaohong,  
; APPLICANT: Malyankar, Uriel,  
; APPLICANT: Miller, Charles,  
; APPLICANT: Millet, Isabelle,  
; APPLICANT: Patturajan, Meera,  
; APPLICANT: Rothenberg, Mark,  
; APPLICANT: Sciore, Paul,  
; APPLICANT: Shenoy, Suresh,  
; APPLICANT: Shimkets, Richard,  
; APPLICANT: Si, Jingsheng,  
; APPLICANT: Smithson, Glenda,  
; APPLICANT: Spytek, Kimberly,  
; APPLICANT: Stone, David,  
; APPLICANT: Taupier, Raymond, Jr.,  
; APPLICANT: Tchernev, Velizar,  
; APPLICANT: Vernet, Corine,  
; APPLICANT: Zerhusen, Brian  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF  
; FILE REFERENCE: 21402-222CIP  
; CURRENT APPLICATION NUMBER: US/10/107,782  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR FILING DATE: 2001-04-20  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 215  
; SOFTWARE: CuraSeglist version 0.1  
; SEQ ID NO 103  
; LENGTH: 1961  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-107-782-103

Query Match 6.8%; Score 102.5; DB 15; Length 1961;  
Best Local Similarity 19.5%; Pred. No. 42;  
Matches 60; Conservative 47; Mismatches 114; Indels 87; Gaps 11;  
QY 7 EQTVVVKSAIE----TADGALDLYNKYLDQVVPWKTFDETIKE-----LSRFRKQEYS 55  
Db 1166 EQEWNILKTTLEBEAKTHAQIQEMRQKHSQAV--BELASQLEQTKRVKANLEKAKQTL 1223  
QY 56 QEASVLVDIKVLLMSQDKYFEATQTVYEWGVTQLLSAYILLFDEYNEKKASAKQDI 115

Db 1224 NERGELANEVKKVLLQGRD-----SEHKRKKVEAQLOE 1256  
QY 116 LTRILDDGVKKLNEAKSLTSSQSFNNASGKLALD---SOLTNDPSEKSSYFQSOVDR 172  
Db 1257 LQVKNEGEVRTELADKVTQLQVELDQVLTGLSSQSDSKSLTKDPSALESQLDQTEL 1316  
QY 173 IRKEAYAGAAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTVQ---NFTSLSAT 228  
Db 1317 LQEN-----ROKL--SLSTKLQKQVEDEKNSFRELEEE 1348  
QY 229 VKQANKDIDAAKLKLATEIAAIGEIKETETTRFYVDYDDL-MLSLLKGAAKMINTCNE 287  
Db 1349 EBEAKHNLEK--QIATLHAQVADMKKME-----DSVGCLETAEEVKRLKQDLEG 1397  
QY 288 YQQRHGKK 295  
Db 1398 LSQRHEEK 1405

## RESULT 48

US-10-369-493-2102  
; Sequence 2102, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkie, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 2102  
; LENGTH: 451  
; TYPE: PRT  
; ORGANISM: Schizosaccharomyces pombe  
US-10-369-493-2102

Query Match 6.7%; Score 102; DB 14; Length 451;  
Best Local Similarity 21.6%; Pred. No. 6.4;  
Matches 81; Conservative 45; Mismatches 131; Indels 118; Gaps 15;  
QY 7 EQTVVVKSA-----LETADGALDLYNKYLDQVVPWKTFDETIKELSRFKOEY--- 54  
Db 4 EENVKEAKNAFNLTQTLSDVEDRDDAL-----DKIVEELRIKKSEVLAAAN 47  
QY 55 --SQEASVLVGD-----IKVLLMSQDKYFEATQTVYEW-----CGVVTQLLSAYI 98  
Db 48 AEDMKAALKLAESGKLSSSMWKELDLSSSDKYSVMQGVLDVKSLLPDLGRVTVARS--- 104  
QY 99 LFFDEYNEKKAS-----AQXDIILIRLDGKKLN-----BAQSKLLTSSQ 139  
Db 105 -LDDGLDLYKVCVPGVGLLVIFEARPEVIINITSLAISKGNVAVLVKGGTSAKSAFALS 163  
QY 140 SFNNASGK-----LLALDSQLTNDFSEKSSYFQSOVDRKEAYAG 180  
Db 164 VRSALGKSKVPOANVOLQVSRREVSQLLKLDYIDLVIIPRGSTNLVHRHKDNTKIIPVLG 223  
QY 181 AAAGI-----VAGPFGI--IISYSIAAGVIEGKLIPE--LNNRLKTVQNFPTS 224  
Db 224 HAAGLCSMYVHEDADDELASKVLVDGKTDYPAACNAIETILLINEAVLSHLPKIAETL 283  
QY 225 LSATVQANKDIDAALKLATEIAAIGEIKETETTRFYVDYDDLMLSLK---GAACK 280  
Db 284 AKVTLK-----CDPASLKVLDKMPKVSALVPEPSVDQDYNTEFSDLIILAIKTVP 338  
QY 281 MINTCEYQQRHGKK 295

Db 339 HINT-----HGSK 346

RESULT 49

US-10-369-493-3210

Sequence 3210, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 3210

LENGTH: 1015

TYPE: PRT

ORGANISM: Neurospora crassa

US-10-369-493-3210

Query Match

Best Local Similarity 20.5%; Pred. No. 19;

Mismatches 65; Conservative 57; Mismatches 123; Indels 72; Gaps 12;

Matches 65; Conservative 57; Mismatches 123; Indels 72; Gaps 12;

QY 18 ETADGALDLYNKYL-DQVIPKTFDETIKELSRPKQVQSASVLDGDIKVL-----LM 70

Db 377 EDPASLADQKSLQDQDRSEKKNALNQKITNLKNSIEBANVALKMKLMQVERDSL 436

QY 71 DSQDK-----YFETATQVVEWCGVVTQLLSAYILLFD-----EYNEKKAQAKDI 115

Db 437 DIQQRQEGDIKSLNQVLDLQKLSKASAGADLKDIOSLRLKNSLFDQQRVQAEKV 496

QY 116 LRIILDDGVKLINEAKSLITSSQFNNSAGKLLALDSQLTNDSEKSSYFQSOVDRIK 175

Db 497 LQQLSQTARLEKVEITL-----KNTASQKMDLETQ-RNEWSKAKKVLREISRLKK 548

QY 176 EAVAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELNNRLK-----TVQN 220

Db 549 EV-----DLARKNSAAPTIVEIKPEPNIAESITRVEQLMKEQKPLPLTARD 595

QY 221 FFTLSATVK--QANKIDDAKKLATEIAAIGEIKTETETTRFYVDYDDLMLSLKGA 278

Db 596 FVQGLVAKLKDAQGHKDLDAEDVIRECIGLLNRLKEP-----VFDEKDKEMLQTLQ--- 648

QY 279 KMINTCNEYQQRHGKK 295

Db 649 ----NELNE--ALRKGKE 660

LENGTH: 1038

TYPE: PRT

ORGANISM: Candida albicans

US-10-032-585-7776

Query Match

Best Local Similarity 20.8%; Pred. No. 20;

Mismatches 64; Conservative 59; Mismatches 118; Indels 66; Gaps 13;

Matches 64; Conservative 59; Mismatches 118; Indels 66; Gaps 13;

QY 7 BOTVEVVKSS---ALETADGALDLYNKYLQDVIPKTFDETIKELSRPKQV-QSASV 62

Db 237 EETIQALKSERNELTAKVSELEDYMKHSE-----VEPDVVMKQNDPEQERIHELEA 291

QY 63 GDIKVLMDSQDKYFETATQVVEWCGVVTQLLSAYILLFDYNEKKAQAKDILRLD- 121

Db 292 LHQTEATIQQSOSRENTE-----LQLOKLTSLTELDKQOEMNRLKASKNENLEMDLSEK 344

QY 122 -DGVKLINEAKSLITSSQFNNSAGKLLALDSQLTN--DFSEK-SSYFQSOVDRIK-KE 176

Db 345 TDNLKELN---NKVLSQAEINLLETKLDLTNSQFENNTDGNKLMKNLESQNKVQTOE 401

QY 177 AVAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELNNRLKTVQNPFTLSATVQANKDI 236

Db 402 AF-----IDELHHEQKTDNEY-----KAKIKDL 425

QY 237 DAAKKLATEIAAIGEIKTETETTRFYVDYDDLML--SLKGAAKKMINTCNEYQQR--- 291

Db 426 EYENAOQLSEISIRAKNSQYDPEAQHYEIDQLQENAOQLKDNVKNYLNFKELKDKVE 485

QY 292 HGKKTIF 298

Db 486 HAHQIAF 492

Search completed: January 5, 2005, 11:08:47

Job time : 66.7056 secs

RESULT 50

US-10-032-585-7776

Sequence 7776, Application US/10032585

Publication No. US20030180953A1

GENERAL INFORMATION:

APPLICANT: Terry, Roemer D.

APPLICANT: Bo, Jiang

APPLICANT: Charles, Boone

APPLICANT: Howard, Bussey

TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery

FILE REFERENCE: 10182-005-999

CURRENT APPLICATION NUMBER: US/10/032,585

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 8000

SOFTWARE: PatentIn version 3.1

SEQ ID NO 7776

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:43:53 ; Search time 55.6826 Seconds  
(without alignments)  
3151.602 Million cell updates/sec

Title: US-09-993-292B-2

Perfect score: 1515

Sequence: 1 MTSIFAEQTVVVKSAIETA.....NEYQQRHGKTLFEVDPVAS 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : UniProt 02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Match | Length | ID           | Description          |
|------------|-------|-------|--------|--------------|----------------------|
| 1          | 1498  | 98.9  | 302    | 1 HLYE_SALTI | Q8727 salmonella     |
| 2          | 1461  | 96.4  | 302    | 1 HLYE_SALPA | Q9316 salmonella     |
| 3          | 1397  | 92.2  | 302    | 1 HLYE_ECOLI | P77335 escherichia   |
| 4          | 1382  | 91.2  | 302    | 1 HLYE_ECO57 | Q9reb3 escherichia   |
| 5          | 1110  | 73.3  | 300    | 2 Q9X2S8     | Q9x288 escherichia   |
| 6          | 519   | 34.3  | 113    | 1 HLEL_SHIFL | Q9rct3 shigella fl   |
| 7          | 395   | 26.1  | 93     | 1 HLEL_ECOL6 | Q8fi27 escherichia   |
| 8          | 130.5 | 8.6   | 895    | 2 Q9LIW7     | Q9liw7 oryza sativ   |
| 9          | 128.5 | 8.5   | 4007   | 2 Q7SH24     | Q7shx4 neurospora    |
| 10         | 126   | 8.3   | 1521   | 2 Q6G015     | Q6g015 bartonella    |
| 11         | 123.5 | 8.2   | 713    | 2 Q6MT03     | Q6mt03 mycoplasma    |
| 12         | 123.5 | 8.2   | 713    | 2 CAE77235   | Caet77235 mycoplasma |
| 13         | 122.5 | 8.1   | 693    | 2 Q7M918     | Q7m918 wolpinella s  |
| 14         | 122   | 8.1   | 495    | 2 P71497     | P71497 mycoplasma    |
| 15         | 121   | 8.0   | 2723   | 2 Q7ROB6     | Q7rob6 plasmodium    |
| 16         | 119   | 7.9   | 541    | 2 Q74DE2     | Q74de2 geobacter s   |
| 17         | 119   | 7.9   | 541    | 2 AAR34750   | Aar34750 geobacter   |
| 18         | 118   | 7.8   | 1496   | 2 Q9SZK7     | Q9szk7 arabidopsis   |
| 19         | 118   | 7.8   | 1877   | 2 Q6BFD6     | Q6bfd6 parametium    |
| 20         | 117.5 | 7.8   | 1274   | 2 Q7NSU5     | Q7nsu5 photorhabd    |
| 21         | 117.5 | 7.8   | 1224   | 2 Q7QQ04     | Q7qq04 giardia lam   |
| 22         | 116   | 7.7   | 1023   | 1 HLY1_ECOL1 | P09983 escherichia   |
| 23         | 116   | 7.7   | 2033   | 2 Q7XEH4     | Q7xeh4 oryza sativ   |
| 24         | 115.5 | 7.6   | 622    | 1 SR68_CAEEL | Q20822 caenorhabdi   |
| 25         | 115   | 7.6   | 587    | 2 Q97WH8     | Q97wh8 sulfobolus    |
| 26         | 114.5 | 7.6   | 478    | 2 Q640E7     | Q640e7 bacterioph    |
| 27         | 114.5 | 7.6   | 478    | 2 Q31954     | Q31954 bacillus su   |
| 28         | 114.5 | 7.6   | 588    | 2 Q9VGA8     | Q9vga8 drosophila    |
| 29         | 114.5 | 7.6   | 652    | 2 Q7CMF0     | Q7cmf0 bacillus an   |
| 30         | 114.5 | 7.6   | 652    | 2 Q9X360     | Q9x360 bacillus an   |
| 31         | 114.5 | 7.6   | 652    | 2 AAT28865   | Aat28865 bacillus    |

#### ALIGNMENTS

#### RESULT 1

| ID | HLYE_SALTI  | STANDARD; | PRT; | 302 AA. |
|----|---|-----------|------|---------|
| AC | Q8Z727; Q934C4;   |           |      |         |
| DT | 29-MAR-2004 (Rel. 43, Created)  |           |      |         |
| DT | 29-MAR-2004 (Rel. 43, Last sequence update)   |           |      |         |
| DT | 01-OCT-2004 (Rel. 45, Last annotation update)   |           |      |         |
| DE | Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA)  |           |      |         |
| GN | Names-hlyE; Synonyms=clyA, sheA; OrderedLocusNames=STY1498, t1477;  |           |      |         |
| OS | Salmonella typhi.   |           |      |         |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;   |           |      |         |
| OC | Enterobacteriaceae; Salmonella.   |           |      |         |
| OX | NCBI_TaxID=601;   |           |      |         |
| RN | [1]   |           |      |         |
| RP | SEQUENCE FROM N.A.  |           |      |         |
| RC | STRAIN=Ty21a, and SMI S2369/96;   |           |      |         |
| RX | MEDLINE=2215712; PubMed=12228306;   |           |      |         |
| RA | Oscarsson J., Westermarck M., Loeffdahl S., Olsen B., Palmgren H., Mizunoe Y., Wai S.N., Uhlin B.E.;  |           |      |         |
| RA | "Characterization of a pore-forming cytotoxin expressed by Salmonella enterica serovar typhi and paratyphi A.";   |           |      |         |
| RL | Infect. Immun. 70:5759-5769(2002).  |           |      |         |
| RN | [2]   |           |      |         |
| RP | SEQUENCE FROM N.A.  |           |      |         |
| RC | STRAIN=CT18.  |           |      |         |
| RX | MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;  |           |      |         |
| RA | Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M., Baker S., Basham D., Brooks K., Chillingworth T., Connerton P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.; |           |      |         |
| RA | "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.";  |           |      |         |
| RL | Nature 413:848-852(2001).   |           |      |         |
| RN | [3]   |           |      |         |
| RP | SEQUENCE FROM N.A.  |           |      |         |
| RC | STRAIN=Ty2 / ATCC 700931;   |           |      |         |
| RX | MEDLINE=22531367; PubMed=12644504;  |           |      |         |
| RA | Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;   |           |      |         |
| RA | "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18.";  |           |      |         |
| RL | J. Bacteriol. 185:2330-2337(2003).  |           |      |         |
| CC | -I- FUNCTION: Toxin, which has some hemolytic activity towards mammalian cells. Acts by forming a pore-like structure upon contact with mammalian cells (By similarity).  |           |      |         |

|    |       |     |      |   |            |                      |
|----|-------|-----|------|---|------------|----------------------|
| 32 | 114.5 | 7.6 | 779  | 2 | Q7RM79     | Q7rm79 plasmodium    |
| 33 | 114.5 | 7.6 | 1885 | 2 | Q869B8     | Q869b8 dictyosteli   |
| 34 | 114   | 7.5 | 1720 | 2 | Q9HGQ7     | Q9hgq7 homo sapien   |
| 35 | 114   | 7.5 | 1189 | 2 | Q8CFI0     | Q8cfi0 staphylococ   |
| 36 | 114   | 7.5 | 1611 | 2 | Q7RD43     | Q7rd43 plasmodium    |
| 37 | 113.5 | 7.5 | 538  | 2 | Q9Z429     | Q9z429 pseudomonas   |
| 38 | 113.5 | 7.5 | 727  | 2 | Q8Z0N2     | Q8z0n2 anabaena sp   |
| 39 | 113.5 | 7.5 | 1127 | 2 | Q9VVT6     | Q9vvt6 melanoplus    |
| 40 | 113.5 | 7.5 | 1363 | 2 | Q874Y4     | Q874y4 podospora a   |
| 41 | 113.5 | 7.5 | 1489 | 2 | Q7RJ32     | Q7rj32 plasmodium    |
| 42 | 113.5 | 7.5 | 1906 | 2 | Q6C359     | Q6c359 yarrowia li   |
| 43 | 113.5 | 7.5 | 3102 | 2 | O45614     | O45614 caenorhabdi   |
| 44 | 113   | 7.5 | 465  | 2 | Q73AJ3     | Q73aj3 bacillus ce   |
| 45 | 113   | 7.5 | 465  | 2 | AAS40708   | Aas40708 bacillus ce |
| 46 | 113   | 7.5 | 956  | 1 | YEF3_YEAST | P32618 saccharomyc   |
| 47 | 113   | 7.5 | 1005 | 1 | RA50_METJA | P58718 methanococ    |
| 48 | 113   | 7.5 | 1628 | 1 | NAGH_CLOPE | P26831 clostridium   |
| 49 | 112.5 | 7.4 | 927  | 2 | Q71X69     | Q71x69 listeria mo   |
| 50 | 112.5 | 7.4 | 927  | 2 | AAT05096   | Aat05096 listeria    |





## RESULT 3

HYLIE ECOLI STANDARD; PRT; 302 AA.  
AC P7735; Q47276; Q8VU70; Q9R3G4;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 29-MAR-2004 (Rel. 43, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Hemolysin E, chromosomal (Hemolysis-inducing protein) (silent  
DE hemolysin sheA) (Cytotoxin ciya) (Latent pore-forming 34 kDa  
DE haemolysin)  
GN Nameshlve; Synonyms=ciya, sheA, hpr; OrderedLocusNames=bll182;  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / XLI-BLUE;  
RA McNamara P.J., Iandolo J.J., Uhlich G.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA del Castillo F.J., Leal S.C., Moreno F., del Castillo I.;  
RT "The Escherichia coli K-12 sheA gene encodes a 34-kDa secreted  
RT haemolysin";  
RL Mol. Microbiol. 25:107-115(1997).  
RN [3]  
RP SEQUENCE FROM N.A., SEQUENCE OF 1-12, SUBCELLULAR LOCATION, AND  
RP INDUCTION.  
RC STRAIN=K12;  
RA Ludwig A., Bauer S., Benz R., Bergmann B., Goebel W.;  
RT Analysis of the siya-controlled expression, subcellular localization  
RT and pore-forming activity of a 34 kDa haemolysin (ciya) from  
RT Escherichia coli K-12";  
RL Mol. Microbiol. 31:557-567(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3030-2;  
RA Xing J., Fernandez S.V., Kapur V., Barletta R.G., Moxley R.A.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map";  
RL DNA Res. 3:137-155(1996).  
RN [7]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CH9802;  
RA Chang G.-N., Ho K.-C.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [8]

RP SEQUENCE OF 1-295 FROM N.A.  
RC STRAIN=K12 / XLI-BLUE;  
RA King C.H., Shinnick T.M.;  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP SEQUENCE OF 1-155 FROM N.A.  
RC STRAIN=K12 / AB1157;  
RA Woodgate R.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
RN [10]  
RP SEQUENCE OF 1-17, AND MUTAGENESIS OF 87-GLY--VAL-89; 142-ASN-ALA-143;  
RP 182-ALA-GLY-183; 186-ALA-GLY-187; ASP-267 AND 292-GLY-LYS-293.  
RX MEDLINE=99316011; PubMed=10383763;  
RA Oscarson J., Mizunoe Y., Li L., Lai X.-H., Wieslander A., Uhlin B.E.;  
RT "Molecular analysis of the cytolytic protein ciya (SheA) from  
RT Escherichia coli";  
RL Mol. Microbiol. 32:1226-1238(1999).  
RN [11]  
RP PARTIAL SEQUENCE, FUNCTION, SUBCELLULAR LOCATION, OLIGOMERIZATION, AND  
RP DISULFIDE BOND FORMATION.  
RX MEDLINE=22894281; PubMed=14532000;  
RA Wai S.N., Lindmark B., Soederblom T., Takade A., Westermark M.,  
RA Oscarson J., Jass J., Richter-Dahlfors A., Mizunoe Y., Uhlin B.E.;  
RT "Vesicle-mediated export and assembly of pore-forming oligomers of the  
RT enterobacterial ciya cytotoxin";  
RL Cell 115:25-35(2003).  
RN [12]  
RP MASS SPECTROMETRY, DISULFIDE BOND, AND MUTAGENESIS OF TYR-96; ASN-156;  
RP TYR-164 AND ARG-260.  
RX MEDLINE=20576278; PubMed=11006277; DOI=10.1074/jbc.M005420200;  
RA Atkins A., Wyborn N.R., Wallace A.J., Stillman T.J., Black L.K.,  
RA Fielding A.B., Hisakado M., Artymluk P.J., Green J.;  
RT "Structure-function relationships of a novel bacterial toxin,  
RT hemolysin E. The role of alpha G";  
RL J. Biol. Chem. 275:41150-41155(2000).  
RN [13]  
RP INDUCTION.  
RX MEDLINE=20507803; PubMed=11053378;  
RA Westermark M., Oscarson J., Mizunoe Y., Urbonaviciene J., Uhlin B.E.;  
RT "Silencing and activation of ciya cytotoxin expression in Escherichia  
RT coli";  
RL J. Bacteriol. 182:6347-6357(2000).  
RN [14]  
RP INDUCTION.  
RX MEDLINE=22053220; PubMed=12057949;  
RA Spory A., Bosserhoff A., von Rhein C., Goebel W., Ludwig A.;  
RT "Differential regulation of multiple proteins of Escherichia coli and  
RT Salmonella enterica serovar Typhimurium by the transcriptional  
RT regulator siya";  
RL J. Bacteriol. 184:3549-3559(2002).  
RN [15]  
RP MUTANT PMWK16 DEL.  
RX MEDLINE=22829856; PubMed=12949101;  
RA Wai S.N., Westermark M., Oscarson J., Jass J., Maier E., Benz R.,  
RA Uhlin B.E.;  
RT "Characterization of dominantly negative mutant ciya cytotoxin  
RT proteins in Escherichia coli";  
RL J. Bacteriol. 185:5491-5499(2003).  
RN [16]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RX MEDLINE=20123445; PubMed=10660049;  
RA Wallace A.J., Stillman T.J., Atkins A., Jamieson S.J., Bullough P.A.,  
RA Green J., Artymluk P.J.;  
RT "E. coli hemolysin E (HlyE, ciya, sheA): X-ray crystal structure of  
RT the toxin and observation of membrane pores by electron microscopy";  
RL Cell 100:265-276(2000).  
CC -1- FUNCTION: Toxin, which has some hemolytic activity towards  
CC mammalian cells. Acts by forming a pore-like structure upon  
CC contact with mammalian cells.  
CC -1- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a  
CC monomer, while in outer membrane vesicles, it oligomerizes to form  
CC a pore structure that is active. Probably forms an octamer.  
CC -1- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer



```

CC -I- SIMILARITY: Belongs to the hemolysin E family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ238954; CAB64962.1; ALT INIT.
CC EMBL; AE055335; AAG56033.1; ALT INIT.
CC EMBL; AP002555; BAB35100.1; ALT_INIT.
CC PIR; E85696; E85696.
CC PIR; E90838; E90838.
CC HSSP; P77335; IQOY.
CC DR InterPro; IPR010356; HlyE.
CC pfam; PF06109; HlyE; 1.
CC KW Complete proteome; Cytolysis; Hemolysis; Toxin; Transmembrane.
CC FT INIT MET 0 By similarity.
CC FT TRANSHEM 182 202 Potential.
CC FT DISULFID 86 284 In monomeric form (By similarity).
CC SQ SEQUENCE 302 AA; 33585 MW; F261E29B1D5FC87 CRC64;

Query Match 91.2%; Score 1382; DB 1; Length 302;
Best Local Similarity 90.1%; Pred. No. 1.8e-84;
Matches 272; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 2 TSIFAEQTVWVKSIAETADGALDLYNKYLDQVTPWKTPTDETIKELSRFKQYEQSASVL 61
DB 1 TEIVADKTVVKNNAIETADGALDLYNKYLDQVTPWKTPTDETIKELSRFKQYEQSASVL 60

QY 62 VGIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRILD 121
DB 61 VGNIKTLLMDSDQKYFEATQTVYEWCGVATQLLAAYILLFDEYNEKKASAKDILIKVLD 120

QY 122 DGVKLENAQKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKRAYAGA 181
DB 121 DGIITLENAQKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKRAYAGA 180

QY 182 AAGVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTQNFFTLSATVKQAKDIDAACL 241
DB 181 AAGVAGPFGLLIISYSIAAGVIEGKLIPELKNKLSQSFVFTLSNTVKQAKDIDAACL 240

QY 242 KLATEIAAIGETETETTRFFVYDDMLSLKGAARKMINTCNEYQQRHGKTLFEVP 301
DB 241 KLATEIAAIGETETETTRFFVYDDMLSLKGAARKMINTCNEYQQRHGKTLFEVP 300

QY 302 DV 303
DB 301 EV 302

RESULT 5
QX288 QX288 PRELIMINARY; PRT; 300 AA.
AC QX288;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hemolysin.
GN Name=hlyE;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=964660;
RX MEDLINE=99242013; PubMed=10227474;
RA Reinhold J., Starr N., Maurer J., Lee M.D.;
RT "Identification of a new Escherichia coli She haemolysin homolog in
RT avian E. coli.";

```

```

RL Vet. Microbiol. 66:125-134(1999).
DR EMBL; AF052225; RAD28079.1; -.
DR HSSP; P77335; IQOY.
DR InterPro; IPR010356; HlyE.
DR pfam; PF06109; HlyE; 1.
SQ SEQUENCE 300 AA; 33555 MW; 6580B66C44A7B4BC CRC64;

Query Match 73.3%; Score 1110; DB 2; Length 300;
Best Local Similarity 74.1%; Pred. No. 2.7e-66;
Matches 215; Conservative 38; Mismatches 37; Indels 0; Gaps 0;

QY 6 AEQTVWVKSIAETADGALDLYNKYLDQVTPWKTPTDETIKELSRFKQYEQSASVLVGI 65
DB 4 ADQTVETVTAIDTAKALDLYNKYLDQVTPWTFNDTVKELSRFKQYEQSASVLVGEI 63

QY 66 KVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRILDGVK 125
DB 64 KSLMNSQDRYPEATQTVYEWCGVVTQLLTAYLSLFNEYDEKKASAKDILIKVLDGII 123

QY 126 KLENAQKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKRAYAGAAGI 185
DB 124 KLEKQOQSLHASSQSFNSAGKLLALDSQLTNDSEKSDYFQSQVDRIKRAYAGAAGV 183

QY 186 VAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTQNFFTLSATVKQAKDIDAACLKAT 245
DB 184 VGRPFGLLIISYSIAAGVIEGKLIPELKEKLSVKDFESLAATVKSANTDIDRAKSLKD 243

QY 246 EIAAIGETETETTRFFVYDDMLSLKGAARKMINTCNEYQQRHGK 295
DB 244 EISVIGDLKTETETTRFFVYDDMLKQLQDSATKLSLSCNEYQQRHGK 293

RESULT 6
HLEL SHIFL
ID HLEL SHIFL STANDARD; PRT; 113 AA.
AC Q9RGT3; Q9RMG1;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemolysin E-like protein.
GN OrderedLocusNames=SF1171, S1259;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12022 / Serotype 2b;
RX MEDLINE=20123445; PubMed=10660049;
RA Wallace A.J., Stillman T.J., Atkins A., Jamieson S.J., Bullough P.A.,
RA Green J., Artymuk P.J.;
RT "E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of
RT the toxin and observation of membrane pores by electron microscopy.";
RL Cell 100:265-276(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CECT 585 / Serotype 2a;
RX MEDLINE=20322319; PubMed=10865950; DOI=10.1016/S0923-2508(00)00143-1;
RA del Castillo F.J., Moreno F., del Castillo I.;
RT "Characterization of the genes encoding the SheA haemolysin in
RT Escherichia coli O157:H7 and Shigella flexneri 2a.";
RL Res. Microbiol. 151:229-230(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";

```

RL Nucleic Acids Res. 30:4432-4441(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RA "Complete genome sequence and comparative genomics of Shigella  
 RT flexneri serotype 2a strain 2457T.";  
 RT Infect. Immun. 71:2775-2786(2003).  
 CC -1- SIMILARITY: Belongs to the hemolysin E family.  
 CC -1- CAUTION: Although it is strongly related to the hemolysin E toxin  
 CC from E.coli K-12 strain, it lacks all the C-terminal part of the  
 CC protein, due to a deletion that creates a frameshift, and it is  
 CC therefore not functional. May be a pseudogene.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AF200955; AAF13995.1; -;  
 DR EMBL; AJ238955; CAB65415.1; ALT INIT.  
 DR EMBL; AE015143; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AE016982; -; NOT\_ANNOTATED\_CDS.  
 DR HSSP; P77335; 100Y.  
 DR InterPro; IPR010356; HlyE.  
 DR Pfam; PF06109; HlyE; 1.  
 DR SEQUENCE 113 AA; 12879 MW; C328908D14C5C4EB CRC64;  
 SQ  
 Query Match 34.3%; Score 519; DB 1; Length 113;  
 Best Local Similarity 91.0%; Pred. No. 2.9e-27;  
 Matches 101; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 MTSIFAQTVVVKSAIETADGALDLYNKYLDQVTPWTFETIKELSRFKQYSQASV 60  
 DB 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPWTFETIKELSRFKQYSQASV 60  
 QY 61 LVGDIKVLWDSQDKYFEATQVYEWCGVVTQLLSAYILLDFEYNEKKASA 111  
 DB 61 LVGDIKTLWDSQDKYFEATQVYEWCGVVTQLLSAYILLDFEYNEKKASA 111  
 RESULT 7  
 HLEL\_ECOL6 STANDARD; PRT; 93 AA.  
 AC QBF127;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Hemolysin E-like protein.  
 GN OrderedLocusNames=cl630;  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OC NCBI\_TaxID=217992;  
 (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6.H1 / CFT073 / ATCC 700928 / UPEC;  
 RX MEDLINE=22382234; PubMed=12471157; DOI=10.1073/pnas.252529799;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J.S., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RA "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 CC -1- SIMILARITY: Belongs to the hemolysin E family.

-1- CAUTION: Although it is strongly related to the hemolysin E toxin  
 CC from E.coli K-12 strain, it lacks all the N-terminal part of the  
 CC protein, and it is therefore probably not functional. May be a  
 CC pseudogene.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AE016759; AAN80095.1; -;  
 DR HSSP; P77335; 100Y.  
 DR InterPro; IPR010356; HlyE.  
 DR Pfam; PF06109; HlyE; 1.  
 DR Complete proteome.  
 DR SEQUENCE 93 AA; 10590 MW; CE1625028DC36DE6 CRC64;  
 SQ  
 Query Match 26.1%; Score 395; DB 1; Length 93;  
 Best Local Similarity 82.8%; Pred. No. 4.6e-19;  
 Matches 77; Conservative 9; Mismatches 7; Indels 0; Gaps 0;  
 QY 211 LNNRLKTVONFTSLSATVKQANKDIDAALKLATEIAALICEIKETETTRFYVDYDLM 270  
 DB 1 MNKLKLSALSPFTLSNTVTKQANKDIDAALKLATEIAALGVIKETETTRFYVDYDLM 60  
 QY 271 LSLKGAARKMINTCNEYQQRHGKTKLFEVPDV 303  
 DB 61 LSLKGAARKMINTCNEYQQRHGKTKLFEIDEV 93  
 RESULT 8  
 Q9LIW7 PRELIMINARY; PRT; 895 AA.  
 ID Q9LIW7  
 AC Q9LIW7  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Similar to an Arabidopsis thaliana chromosome BAC genomic  
 DE sequence.  
 DE Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hsing Y.C., Chow T., Chen C., Wu H., Chu M., Chao Y., Liu S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AP001111; BAA90502.1; -;  
 DR Gramene; Q9LIW7; -;  
 DR SEQUENCE 895 AA; 100520 MW; BAA7BFDF70FPA3 CRC64;  
 SQ  
 Query Match 8.6%; Score 130.5; DB 2; Length 895;  
 Best Local Similarity 22.1%; Pred. No. 3.3;  
 Matches 76; Conservative 66; Mismatches 123; Indels 79; Gaps 17;  
 QY 8 QTVVVKSAIETADGALD-----LYNKYLDQVTPWTFETIK-----ELSRFKQY 54  
 DB 274 EKVEILSSEVVRKGLDSTAESEESKRETELV--KNLESEVSVLKGKLEARIIEER 331  
 QY 55 SQEASVLVGDIKVLLMDSQDKYFEATQVYEW---CGVVTQLLSAYILLDFEY-NEKKAS 110  
 DB 332 LAETKLEELKSEVADAKKAESARQLFEWKHAGLLEMELEA-VTUSDKPKGSSLAS 390  
 QY 111 AQKDILRIILDDGVKKLNEAKQSLITSSQSFNNASGKLALD---SOLTNDPSEKSYF- 166  
 DB 391 T-----TEELGKIQSALQDRESEIEVLKGTITALEIEVALLADVNESNEQFD 438  
 QY 167 -----SQSDVRIRKEAYAGAAAGIVAGPFGLIISYSIAAGV-IEG-----KLIPE 210

```
Db 439 ASQEVFGLQTTIDVLRNKLAEAEAESEA-----LNNEKAANVKIEGLTEENVKLISE 492
Qy 211 LN-----NRLKTVQNFYSLVTKOANKDIDAAKLKLATE-----IAAIGEIKTET 257
Db 493 LNETRDREEKRAVEDLTAALS-----BESDKAKEAHERYLSKEDDHEHALAQIGDLKWL 549
Qy 258 ETR-----FVVDYDMLSLKGAACKMINTCNEYQOR-HGKKT 296
Db 550 KSTKSEYVMDLEANYDITCLRNKVDKLEAEVKNYRECSKET 593

RESULT 9
Q7SHZ4
ID Q7SHZ4 PRELIMINARY; PRT; 4007 AA.
AC Q7SHZ4;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU00658.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins D., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Iankiev P., Pedersen D., Nelson M., Washburne M.,
RA Salitrenikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal K., Kamysellis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krysstofova S., Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmari S.A.,
RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR ENBL; AABX01000001; EAA36562.1; -.
DR InterPro; IPR009638; Rezi.
DR InterPro; IPR003900; KID_repeat.
DR Pfam; PF06818; Fezi; 1.
DR Pfam; PF02524; KID; 5.
DR PROSITE; PS01039; SBP_BACTERIAL_3; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 4007 AA; 453246 MW; 67CDPF6EPF1463612 CRC64;

Query Match 8.5%; Score 128.5; DB 2; Length 4007;
Best Local Similarity 21.3%; Pred. No. 27;
Matches 75; Conservative 65; Mismatches 141; Indels 71; Gaps 14;

Qy 8 QTVVVKSAIETADGAL-----DLNKKYLDVIVPKTFDETIK-----ELSRFKQE 53
Db 1156 QTEILRLKQHQSRGVELESRIATIKYKQDLDELSSRNNTSQDAIKLQHENELANPAK 1215
Qy 54 YSQEASLVGDIKVLMDSDQKYPEATQTVYEWGCVVTVQLLSAVI-----LLFDEYNKKA 109
Db 1216 YEQKKQLAVQHKTEMESLDRHYHEKELATQYQERV-QALSALADKKTALAEYKE-QL 1273
Qy 110 SAQKDILIRLDDGKVKLENAQKSL-----LTSQSFPNASKGLALDSQ- 154
Db 1274 SASKAQLDKLKHGKGVKVDLQAKLKSEVAKVTADYEGNLSELRTKHQGVNVLKVHQD 1333
Qy 155 -----LTNDFSEKSSYFOSQVDRIRKEAYAGAAAGVAGPGLIISYIAAGVIEGKL--- 207
```

```
Db 1334 EIKKLFAGHNEKIRNLEHRINDLKAELKQDRA-----EFDKKKALIEGEVATL 1381
Qy 208 ---IPELNNRLKTVQNFYSL-----SATVQOANKDIDAAKLKLATEIAAIGEIKTET- 259
Db 1382 QGKVDKSSKSLSKAEAFNELKXLAELKADVADKSNLQDKLEELSDLKGGQCKTR 1441
Qy 260 -TRFYVDYDMLSLKLG-----AAKMNIN-TCNEYQORHG--KKTILFEVDP 302
Db 1442 IEDFNQINEKMAQLLKAQNELKASQASLNTTTTTEYDAKTAQLEKSLKEKD 1493

RESULT 10
Q6G015
ID Q6G015 PRELIMINARY; PRT; 1521 AA.
AC Q6G015;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=BQ05330;
OS Bartonella quintana (Rochalimaea quintana).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=803;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Toulouse;
RX PubMed=15210978;
RA Alenmark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardell D.H.,
RA Canback B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huvet M.,
RA La Scola B., Holmberg M., Andersson S.G.E.;
RT "The louse-borne human pathogen Bartonella quintana is a genomic
RT derivative of the zoonotic agent Bartonella henselae.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).
DR ENBL; BX897700; CAP26028.1; -.
DR InterPro; IPR009074; Apolipo_A_E_C3.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR010989; t-snare.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1521 AA; 169742 MW; 12CC93BCB2C97920 CRC64;

Query Match 8.3%; Score 126; DB 2; Length 1521;
Best Local Similarity 18.8%; Pred. No. 13;
Matches 70; Conservative 68; Mismatches 148; Indels 90; Gaps 8;

Qy 3 SIPAEQTVVVKSAIETADGALDLNKK-----YLDQVLPWK-----TFDE 42
Db 927 SVLSEQTAQTVESFTTASHNAQTLINETHITSATAIEEVLNERCNVLHSHMQNLKNGY 986
Qy 43 TIKELSRFKQEQYSQEASVLVGDIKVLLMDSQDKYFEATQTVYEWGCVVTVQLLSAYILLFD 102
Db 987 QLSDSVSNHLEESAKQTATQISGHVEKLTETLVNLNQAQNTTESISHLTQHISEQLSLST 1046
Qy 103 EYNEKASAKQDILIRL-----DDGVKLENAQKSLTSSQSFNNASGK 147
Db 1047 QDAEQRIYAQNESLVNSLTQTNSSETLQTVTAMKEDLVNNISSILKQLNQSIYSPHNSNI 1106
Qy 148 LLALDSQLTNDPFEKSGSYFOSQVDRIRKEAYAGAAAGVAGPGLIISYIAAGVIE--- 204
Db 1107 LLSTVQNDQGFSETANNFRKTNQ-----AAEHLASNAQALNNVVLQGLSQNIF 1158
Qy 205 -----GKLIPELNNRLKTVQNFYSLVTKOANKDIDAAKLKLATEIAAIGEI- 253
Db 1159 EKIGHTSTFGEHAKTSETIHIHLEKSENSLTSTLEBKHKTLKLSALVSKSNEINKLI 1218
Qy 254 -----KTETET-----TRFYVDYDMLSL--LKGRAKK 280
Db 1219 EYENVLSELAFTDTRNTRNTHSLQOQLNQLINEASTRFGAAEDIRRSADERSLSK 1278
Qy 281 MINTCNEYQORHGKKT 296
Db 1279 INNDINESVONLPEKT 1294
```

```

GN CLPB OR MSC_0613.
OS Mycoplasma mycoides (subsp. mycoides SC).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=44101;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PGI;
RC PubMed=14762060;
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundberg J.,
RA Johansson K.-E., Pettersson B., Uhlen M.;
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type
RT strain PGI1, the causative agent of contagious bovine pleuropneumonia
RT (CBPP).";
RL Genome Res. 14:221-227(2004).
DR EMBL; BX842644; CAE77235.1; -.
KW Protease.
SQ SEQUENCE 713 AA; 80782 MW; 1B5D204A9E29BE50 CRC64;

Query Match      8.2%; Score 123.5; DB 2; Length 713;
Best Local Similarity 23.3%; Pred. No. 7.5; Indels 95; Gaps 17;
Matches 79; Conservative 47; Mismatches 118;

QY 2 TSIFAEQTVVVKSAIETADGALDLYNKYLDQVLPWKTFDETI-----KELSRF 50
Db 396 TGLVDRLLISSEKEKLNLEDLLKKYKGDQAI--KAVTSAIMRSRGIKNPKPIGSF 453
QY 51 -----KQEYSQ--ASVLVGDIKVLLMDSQDKYFEATQTV-----YEWCGV 89
Db 454 LFLGPTGVGKTEVARSLADILFNSPKMIRLDMSEYMEKHSVAKLIGAPPVGYEGGR 513
QY 90 VTQLL--SAY--ILLFDEYNEKASAKDILIRILDDG-----VKKLE 129
Db 514 LTEAVRNPSYIVLFDEI--EKAHTDVFNILLQILDDGRLTDSLGTIDFKNTIIVMTSNI 572
QY 130 AQKSLTSSQSFNNASGKLLALDSQLTNDFSEKSYFQSQVDRIKKEAYAGAAAGIVAGP 189
Db 573 ASQYLLTSDE-----LVQVDDQ-----KIQELNKKVFPFLNIDNIVY-- 612
QY 190 FGLIISYSIAAGVIEGKLIPELNNRLKTQVNFSTLS--ATVKQAKDID-----AAKLK- 242
Db 613 FNALSQVTI--GEIVDKVLEELSTRLOEQNYFINFSEARNKIINEGYDRFLFGARPIKR 670
QY 243 -----LATEIAAIGEIKTETETTFYVYDDMLSL 273
Db 671 YIEKNIETLIAHYIISGEV---VENTRYLIDVKNNOFTL 706

RESULT 13
Q7M918 PRELIMINARY; PRT; 693 AA.
ID Q7M918
AC Q7M918;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PUTATIVE MEMBRANE PROTEIN.
GN OrderedLocusNames=WS0888;
OS Wolinella succinogenes.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Wolinella.
OX NCBI_TaxID=844;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DSMZ 1740;
RC MEDLINE=22882897; PubMed=14500908;
RA Baar C., Eppinger M., Radatz G., Simon J., Lanz C., Klimmek O., B.,
RA Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,
RA Meyer F., Lederer H., Schuster S.C.;
RT "Complete genome sequence and analysis of Wolinella succinogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695(2003).
KW Complete proteome.
SQ SEQUENCE 693 AA; 77641 MW; CB57A1F23CFACC2 CRC64;

Query Match      8.2%; Score 123.5; DB 2; Length 713;
Best Local Similarity 23.3%; Pred. No. 7.5;
Matches 79; Conservative 47; Mismatches 118; Indels 95; Gaps 17;

QY 2 TSIFAEQTVVVKSAIETADGALDLYNKYLDQVLPWKTFDETI-----KELSRF 50
Db 396 TGLVDRLLISSEKEKLNLEDLLKKYKGDQAI--KAVTSAIMRSRGIKNPKPIGSF 453
QY 51 -----KQEYSQ--ASVLVGDIKVLLMDSQDKYFEATQTV-----YEWCGV 89
Db 454 LFLGPTGVGKTEVARSLADILFNSPKMIRLDMSEYMEKHSVAKLIGAPPVGYEGGR 513
QY 90 VTQLL--SAY--ILLFDEYNEKASAKDILIRILDDG-----VKKLE 129
Db 514 LTEAVRNPSYIVLFDEI--EKAHTDVFNILLQILDDGRLTDSLGTIDFKNTIIVMTSNI 572
QY 130 AQKSLTSSQSFNNASGKLLALDSQLTNDFSEKSYFQSQVDRIKKEAYAGAAAGIVAGP 189
Db 573 ASQYLLTSDE-----LVQVDDQ-----KIQELNKKVFPFLNIDNIVY-- 612
QY 190 FGLIISYSIAAGVIEGKLIPELNNRLKTQVNFSTLS--ATVKQAKDID-----AAKLK- 242
Db 613 FNALSQVTI--GEIVDKVLEELSTRLOEQNYFINFSEARNKIINEGYDRFLFGARPIKR 670
QY 243 -----LATEIAAIGEIKTETETTFYVYDDMLSL 273
Db 671 YIEKNIETLIAHYIISGEV---VENTRYLIDVKNNOFTL 706

RESULT 12
CAE77235 PRELIMINARY; PRT; 713 AA.
ID CAE77235
AC CAE77235;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 13-APR-2004 (TrEMBLrel. 27, Last annotation update)
DE ATP dependant protease ClpB.

```

```

Query Match      8.1%; Score 122.5; DB 2; Length 693;
Best Local Similarity 21.7%; Pred. No. 8.4; Mismatches 115; Indels 77; Gaps 10;
Matches 67; Conservative 50;

Qy 11 EVKSAIETADGALDLYNKYLDQVPIPKTFDETIKELSRFKQYSQBSASVLGVGDKVLLM 70
Db 316 BEVTSSTKEAQESLTSQRWLQERIQ-----SDISLAQHPTEHLGLSTFEKALVV 367
Qy 71 DSQKYPEATQTVYEWCGVVTOLLISAVILLPDEYNEKKAQKDILRIIDDGVKKLNEA 130
Db 368 SLRQLGEGANEVLISLLQKSSQESSAFLL----EQERMAKESLTVLPAKLKERFERLSEA 423
Qy 131 QKSLLTSS-----QSFNASGKLLALDSQLTNDSEKSSYFQ 167
Db 424 ---FLTNKQGALEGMALDFRGFWNEYATRWRESSELTOGTLLTNQQRASFAELSGVL 480
Qy 168 SQVDRIRKEAYAGAAAGVAGPF-GLIISYSIAAGVIEG--KLIPELN-----NRLKTVO 219
Db 481 AQNSAMGEQIKAG-FEGMTQAGLSGLMLAKALEGGIDGVKKSVMNQTLAKETLQTLQ 539
Qy 220 NFFTSLSATVKQANKIDAAKLIKLAETIAAIGEIKTETETTRFVYDDLMLSLLKGAOK 279
Db 540 NHSQS-----QLGLLGESTTAIQR-----LLELEGAOK 569
Qy 280 KMINTCNEY 288
Db 570 SLKNLALEY 578

RESULT 14
P71497 PRELIMINARY; PRT; 495 AA.
ID P71497
AC P71497;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ClpB (Fragment).
GN Name=clpB;
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OC NCBI_TaxID=2095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25416;
RX MEDLINE=97148974; PubMed=8995799;
RA Falah M., Gupta R.S.;
RT "Phylogenetic analysis of mycoplasmas based on Hsp70 sequences: cloning of the dnaK (hsp70) gene region of Mycoplasma capricolum.";
RL Int. J. Syst. Bacteriol. 47:38-45(1997).
DR EMBL; U51235; AAB09427.1; -.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR001270; Chaperin clpA/B.
DR PRINTS; PR00300; CLPPTPTEASEA.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00871; CLPAB_2; 1.
KW ATP-binding.
FT NON TPR
SQ SEQUENCE 495 AA; 56358 MW; A194DD51FACDF8D0 CRC64;

Query Match      8.1%; Score 122; DB 2; Length 495;
Best Local Similarity 24.6%; Pred. No. 6.1;
Matches 82; Conservative 37; Mismatches 113; Indels 102; Gaps 16;

Qy 2 TSIAEQTVVVKSAIETADGALDLYNKYLDQVPIPKTFDETIKELSR 49
Db 178 TGIIVDKLISSEKRLNLDLLKKYKQGQQAIAVTSIAIMRSRSGKPKDPSGFLF 237
Qy 50 F-----KQEYSQBS-ASVLVGDIKVLLMDSQDKYFEATQTV-----YEWCGVWT 91
Db 238 FGPTGVKTEVARSLADILFNSPKMIRLDMSEYNEKHSVAKLIGAPPYGVYEGGRLT 297

Query Match      8.0%; Score 121; DB 2; Length 2723;
Best Local Similarity 19.4%; Pred. No. 54;
Matches 60; Conservative 54; Mismatches 102; Indels 94; Gaps 9;

Qy 11 EVKSAIETADGALDLYNKYLDQVPIPKTFDETIKELSRFKQE----- 53
Db 1464 EELKGYIDSKSGYKDEVNKNT-----KTIENNKKELEKYKNEAINLNLFSEPMLNKI 1517
Qy 54 --YSQASVLVGDIKVLLMDSQDKYFEATQTVYEWCGVVTOLLISAVILLPDEYNEKKA 111
Db 1518 IQTKDLEIITINIKKWHKES-----ILOADKSEQKINT 1551
Qy 112 QKDILIRILDDGVKKLNEAOKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVD 171
Db 1552 IKNEIQIKDQTTSN-DKSNKAIIGIQASLDKPEFKI-----NDIRTKSNFLKETE 1605
Qy 172 RIRKEAYAGAAAGVAGPFGLLIISYSIAAGVIEGKLIPELNRLKTVQNTFTLSATVKQ 231

```



```

Db 1606 NIEQK-----ISNLSINSQEIK---LKNEDILNTLQKFFSLKDKQKN 1646
Qy 232 -----ANKDIDAAKKLATIAAIGIKETETETFRFYVDYDMLMLSLKG 276
Db 1647 IEDQKTELDNFDSEIENIESDVQKNKYEMGLIEKIKENADTNKQIESTK--ELIKP 1703

Qy 277 AAKKWINTCN 286
Db 1704 TIENLISSEN 1713

RESULT 16
Q74DE2 PRELIMINARY; PRT; 541 AA.
AC Q74DE2;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN Name=hylB; ORFNames=GSU1374;
OS Geobacter sulfurreducens.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC Geobacteraceae; Geobacter.
OX NCBI_TaxID=35554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCA / ATCC 51573;
RX PubMed=14671304;
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,
Heidelberg J.F., Wu D., Wu M., Ward N.L., Beanan M.J., Dodson R.J.,
Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,
Davidson T.M., Zafar H.M., Feldblyum T.V., Utterback T.R.,
Van Aken S.E., Lovley D.R., Fraser C.M.;
RA "Genome of Geobacter sulfurreducens: metal reduction in subsurface
environments.";
RT Science 302:1967-1969(2003).
RL ENBL; AB017190; AAR34750.1; -.
DR TIGR; GSU1374; -.
SQ SEQUENCE 541 AA; 57093 MW; 3C9BEDAEAF7C3812 CRC64;

Query Match 7.9%; Score 119; DB 2; Length 541;
Best Local Similarity 23.5%; Pred. No. 11;
Matches 69; Conservative 44; Mismatches 103; Indels 78; Gaps 12;

Qy 19 TADGALDLYNKYLDQVTPWKTDETIKLSRQKQYSEASVLVGDIVKLLMDSQDKYFE 78
Db 36 TANNGLD--TVYRDRVLP-----LKLKLIADMY-----AVNIVDVSHKVRNGNITWE 82
Qy 79 ATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKQDILIRILDGKVLNEAKSLLTSS 138
Db 83 GRKSVEEAKKTIAEKIQLAYLATNLAEEKKHLEAKPLIKVADATLER-----LASI 134
Qy 139 QSFNNSGKLLALDSQLTNDSEKSYFQSQVD---RIRKEAY-----A 179
Db 135 LSAEDAEALTAFTVSELYPAIDPVSAKFSLSVDDQLKIAKQYDHSGLYRASRTISLVA 194
Qy 180 GAAAGIVAGPGLIISYIAAGVIEGKLIPELNNRLK-----TVQ----- 219
Db 195 IIVGVLIAGTAGLITTSITGPLAEG---SATVQKANKIDIAKKLAT---EIAA-IGEIKTETE 258
Qy 220 -NFTSL-----SATVQKANKIDIAKKLAT---EIAA-IGEIKTETE 258
Db 252 GNMVTSRLHLIAEAISSHGIASNSQLHATSEQIATGSEVASQVGAVATASE 305

RESULT 17
AAR34750 PRELIMINARY; PRT; 541 AA.
ID AAR34750;
AC AAR34750;
DT 02-MAR-2004 (TRENBLrel. 27, Created)
DT 02-MAR-2004 (TRENBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TRENBLrel. 27, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN HYL B OR GSU1374.
OS Geobacter sulfurreducens.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC Geobacteraceae; Geobacter.
OX NCBI_TaxID=35554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCA / ATCC 51573;
RX PubMed=14671304;
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,
Heidelberg J.F., Wu D., Wu M., Ward N.L., Beanan M.J., Dodson R.J.,
Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,
Davidson T.M., Zafar H.M., Feldblyum T.V., Utterback T.R.,
Van Aken S.E., Lovley D.R., Fraser C.M.;
RA "Genome of Geobacter sulfurreducens: metal reduction in subsurface
environments.";
RT Science 302:1967-1969(2003).
RL ENBL; AB017211; AAR34750.1; -.
DR TIGR; GSU1374; -.
SQ SEQUENCE 541 AA; 57093 MW; 3C9BEDAEAF7C3812 CRC64;

Query Match 7.9%; Score 119; DB 2; Length 541;
Best Local Similarity 23.5%; Pred. No. 11;
Matches 69; Conservative 44; Mismatches 103; Indels 78; Gaps 12;

Qy 19 TADGALDLYNKYLDQVTPWKTDETIKLSRQKQYSEASVLVGDIVKLLMDSQDKYFE 78
Db 36 TANNGLD--TVYRDRVLP-----LKLKLIADMY-----AVNIVDVSHKVRNGNITWE 82
Qy 79 ATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKQDILIRILDGKVLNEAKSLLTSS 138
Db 83 GRKSVEEAKKTIAEKIQLAYLATNLAEEKKHLEAKPLIKVADATLER-----LASI 134
Qy 139 QSFNNSGKLLALDSQLTNDSEKSYFQSQVD---RIRKEAY-----A 179
Db 135 LSAEDAEALTAFTVSELYPAIDPVSAKFSLSVDDQLKIAKQYDHSGLYRASRTISLVA 194
Qy 180 GAAAGIVAGPGLIISYIAAGVIEGKLIPELNNRLK-----TVQ----- 219
Db 195 IIVGVLIAGTAGLITTSITGPLAEG---SATVQKANKIDIAKKLAT---EIAA-IGEIKTETE 258
Qy 220 -NFTSL-----SATVQKANKIDIAKKLAT---EIAA-IGEIKTETE 258
Db 252 GNMVTSRLHLIAEAISSHGIASNSQLHATSEQIATGSEVASQVGAVATASE 305

RESULT 18
Q9SZK7 PRELIMINARY; PRT; 1496 AA.
ID Q9SZK7;
AC Q9SZK7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Hypothetical protein AT4g38070.
DE Name=F20D10.190; Synonyms=AT4g38070;
GN

```

```

Db 1606 NIEQK-----ISNLSINSQEIK---LKNEDILNTLQKFFSLKDKQKN 1646
Qy 232 -----ANKDIDAAKKLATIAAIGIKETETETFRFYVDYDMLMLSLKG 276
Db 1647 IEDQKTELDNFDSEIENIESDVQKNKYEMGLIEKIKENADTNKQIESTK--ELIKP 1703

Qy 277 AAKKWINTCN 286
Db 1704 TIENLISSEN 1713

RESULT 16
Q74DE2 PRELIMINARY; PRT; 541 AA.
AC Q74DE2;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN Name=hylB; ORFNames=GSU1374;
OS Geobacter sulfurreducens.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC Geobacteraceae; Geobacter.
OX NCBI_TaxID=35554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCA / ATCC 51573;
RX PubMed=14671304; DOI=10.1126/science.1088727;
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,
Heidelberg J.F., Wu D., Wu M., Ward N.L., Beanan M.J., Dodson R.J.,
Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,
Davidson T.M., Zafar H.M., Feldblyum T.V., Utterback T.R.,
Van Aken S.E., Lovley D.R., Fraser C.M.;
RA "Genome of Geobacter sulfurreducens: metal reduction in subsurface
environments.";
RT Science 302:1967-1969(2003).
RL ENBL; AB017190; AAR34750.1; -.
DR TIGR; GSU1374; -.
RP InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me Chemotaxis.
DR InterPro; IPR005829; Sug transporter.
DR InterPro; IPR010989; t-snare.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPeignal; 1.
DR PRINTS; PR0260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS00885; HAMP; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN 1.
SQ SEQUENCE 541 AA; 57093 MW; 3C9BEDAEAF7C3812 CRC64;

Query Match 7.9%; Score 119; DB 2; Length 541;
Best Local Similarity 23.5%; Pred. No. 11;
Matches 69; Conservative 44; Mismatches 103; Indels 78; Gaps 12;

Qy 19 TADGALDLYNKYLDQVTPWKTDETIKLSRQKQYSEASVLVGDIVKLLMDSQDKYFE 78
Db 36 TANNGLD--TVYRDRVLP-----LKLKLIADMY-----AVNIVDVSHKVRNGNITWE 82
Qy 79 ATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKQDILIRILDGKVLNEAKSLLTSS 138
Db 83 GRKSVEEAKKTIAEKIQLAYLATNLAEEKKHLEAKPLIKVADATLER-----LASI 134
Qy 139 QSFNNSGKLLALDSQLTNDSEKSYFQSQVD---RIRKEAY-----A 179
Db 135 LSAEDAEALTAFTVSELYPAIDPVSAKFSLSVDDQLKIAKQYDHSGLYRASRTISLVA 194
Qy 180 GAAAGIVAGPGLIISYIAAGVIEGKLIPELNNRLK-----TVQ----- 219

```



OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 RN NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Wedler H., Kutzner M., Wambutt R., Bancroft I., Mewes H.W.,  
 RA Mayer K.F.X., Schueller C.;  
 RA Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Wedler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL035538; CAB37547.1; -  
 DR EMBL; AL161592; CAB80472.1; -  
 DR PIR; T05634; T05634.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR Pfam; PF00010; HLH; 1.  
 DR SMART; SM00353; HLH; 1.  
 DR PROSITE; PS00888; HLH; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1496 AA; 173346 MW; 03BBB75DC2B03FE CRC64;  
 Query Match 7.8%; Score 118; DB 2; Length 1496;  
 Best Local Similarity 19.6%; Pred. No. 42;  
 Matches 64; Conservative 67; Mismatches 122; Indels 74; Gaps 13;  
 QY 4 IFRAQTEVVVKSATETADGALDLYNK--YLDQVTPWKT-----PDETIKLSRP----KQ 52  
 DB 153 IEAEKTVKGMKRGDRDDVVVMBEESQVBEKLNKKEQFKHLEAYEKLKMLFKDSKK 212  
 QY 53 EYSQAEASVLVGDIVL--LMSQDKYFEATQTVVEMC-GVVTQLLSAYILLFDEYNKKA 109  
 DB 213 EWEKSKLDEIYSLQTKDVSRTISEDQKQLQMCNGALQ-----BETRKHLL 263  
 QY 110 SAQKDILIRLDGKVKLNLEAQSLLTSFSSFNASGKLLALDSQLTDFSEKSSYFQ-- 167  
 DB 264 EIQVSEPKAKYDAFAECQDARTQL-----DDLACKRDWEVAELRQTLQSMKDAYPKEM 316  
 QY 168 ----SQVDRIRKEAYAGAAAGIVAGPGLIISYIAAGVIEG---KLIPELNRLKTVQN 220  
 DB 317 KYENGKLEQENRELLGSLK-----ELQBATIQGSGNSALSCLKNKNFRLEN 362  
 QY 221 FFTSLSAT-----VKQAKNDIDAAKLATETIAAIGETKETE-----TTRFYV 264  
 DB 363 IHRKCSANLAKESKEWESQVKEWVEEINDYKLQSKKEALKKEVELENCRSSTAKMRL 422  
 QY 265 DYDDLMLSLKGAQKMNCTNEYQQR 291  
 DB 423 QYBEISIMFL-----VLSRTVSEASQR 444  
 RESULT 19  
 Q6BFD6  
 ID Q6BFD6 PRELIMINARY; PRT; 1877 AA.  
 AC Q6BFD6;  
 DT 01-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=PTMB.434;  
 OS Paramesidium tetraurelia.  
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;

OC Paramesidium.  
 OX NCBI\_TaxID=5888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Stock d4-2;  
 RA Zagulski M., Nowak J.K., Le Mouel A., Nowacki M., Migdalski A.,  
 RA Gromadka R., Noel B., Blanc I., Dessen P., Wincker P., Kellier A.M.,  
 RA Cohen J., Meyer E., Sperling L.;  
 RA "high coding density on the largest Paramesidium tetraurelia somatic  
 RT chromosome";  
 RL Curr. Biol. 0:0-0(2004).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Stock d4-2;  
 RA Nowak J.K., Migdalski A., Gromadka R., Zagulski M.;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CR548612; CAH03635.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 1877 AA; 220626 MW; F68260FE840D2796 CRC64;  
 Query Match 7.8%; Score 118; DB 2; Length 1877;  
 Best Local Similarity 20.3%; Pred. No. 55;  
 Matches 66; Conservative 64; Mismatches 135; Indels 60; Gaps 13;  
 QY 7 EQTVEVVKSAIETADGALDLYNKYLDQVTPWKTDETIKLSRFBKQBYSOEASVYVGDIK 66  
 DB 398 EQNKDILQHVANS-----FNKFLQEQANEQNAVFPKLLYNFFMEHDMHVELLIQVVS 450  
 QY 67 VLIMDSQDKYFEATQTVY----EWCQVVTQLLSAYILLFDEYNKKAQKQDILIRL-- 120  
 DB 451 FNLI-----KYVETNHKKYPIVKSCKRLLESISFLLIVVNEVSKVYNTYINFLSILIT 506  
 QY 121 -DGVVK-----LNEAQSKLLTSQSFN-----NASGKL-----ALDSQL 155  
 DB 507 KEDNFEKSTIEFTLLRGAANLMTCLSKFSPEELKQSPGSTLNTLLVMEQLPQDLKQL 566  
 QY 156 TNDPSEKSSYFQSQVDRIK--EAYAGAAAGIVAGPGLIISY-SIAAGVIEGKLIPELN 212  
 DB 567 VGP-AEMSNY-----DDAIEFGQLYQNLSDVLDGDEQIRVFESIKSLIGURFKQ-N 620  
 QY 213 NRLKTVQNFTSLSATVVKQAKNDIDAAKLATETIAAIGETKETEETETTFRFYVDDLMLS 272  
 DB 621 KEVQLLPKWTSTIKIIEKNPSISLLAQSMIEILISEKVD-----IIVENLKL 671  
 QY 273 LLGAQAKMNCTNEYQQRHGKTL 297  
 DB 672 IIBESRTKMLKNOIVEGQDYTKLTL 696  
 RESULT 20  
 Q7N5U5  
 ID Q7N5U5 PRELIMINARY; PRT; 577 AA.  
 AC Q7N5U5;  
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Similar to putative phage protein.  
 GN OrderedLocNames=plui1837;  
 OS Photobacterium luminescens (subsp. laumondii).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Photobacterium.  
 OX NCBI\_TaxID=141679;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TT01;  
 RX MEDLINE=22957627; PubMed=14528314;  
 RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,  
 RA Taurit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,  
 RA Dassa E., Derose R., Deruelle S., Freysinet G., Gaudrault S.,  
 RA Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,  
 RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;  
 RA "The genome sequence of the entomopathogenic bacterium Photobacterium  
 RT luminescens.";

```

RL Nat. Biotechnol. 21:1307-1313 (2003).
DR EMBL; BX571865; CAB14130.1; -.
DR Photolist; plu1837; -.
KW Complete proteome.
SQ SEQUENCE 577 AA; 65287 MW; B7F86EE4B34BEDE1 CRC64;

Query Match 7.8%; Score 117.5; DB 2; Length 577;
Best Local Similarity 19.2%; Pred. No. 15;
Matches 73; Conservative 56; Mismatches 142; Indels 109; Gaps 12;

QY 17 IETADGALDLYNKYLDQV---IPWKTFF-----DETI 44
DB 129 LETPPAYVFLPFIIDQIKSWSSPNSFENLGOYSNWKGPLIKYFTGYLKPHEFDIEENI 188
QY 45 KELSFRKQOEYSQASVLVGDIKVL-----LMSQDK 75
DB 189 YEYSEIKKESAHKTEKQSAVEVINDVTVDSPITLDNNDFIKIQNEIKNELYLDIDYQTR 248
QY 76 YFEATQTVYEWGVVQTLLSAYILLPDEYNEKKA-----SAQKQIL----- 116
DB 249 LYDAQATI---TSNIYDLESQYALAITSAANELEADYKFAVESIPTDYLECPGLCTLHDNS 305
QY 117 ----IRILDDGVKKLINEAKQSLTSSQSFNNASGKLALDSQLTNDPSEKSSYFQSOVDRIKEAYAGAAAGI 185
DB 306 LPNRALLSEKXDSLNEA-----NSIASKIVELKSSL-NSLNEDAQFITTNEIER 353
QY 173 IRKEAYAGAAA---GIVAGPGLIISYSIAGV-----IEGKLIPELNNRLKTVQNFPTS 224
DB 354 INNKYITDDDAVKKGLIAQVITLSTENVSKNIQIKINDENLNISKANSSIKEKDKQKK 413
QY 225 LSATVKQANKOIDAAKLAKEIATIAAIGIKETETETTRFYVDYDMLSLKGAAGKMI-N 283
DB 414 LUST-KEEELNSSPMKLLANIELALGSTGINSKVSPTDYKQLGGGAEEAARGLLAY 472
QY 284 TCNEYQORHGKKTLPFVPPV 303
DB 473 QLSVLQIIGHAKTCVVPFV 492

RESULT 21
Q7Q004 PRELIMINARY; PRT; 1224 AA.
AC Q7Q004;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE GLP_227_22033_18359.
OS Giardia lamblia ATCC 50803.
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
OX NCBI_TaxID=184922;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WB C6;
RA Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gallin F.D.,
RA Olsen G.J., Sogin M.L.;
RT "Draft sequence of the Giardia lamblia genome.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AACB01000157; EAA37104.1; -.
CC InterPro; IPR002110; ANK.
CC Pfam; PF00023; Ank; 5.
CC PROSITE; PS50088; ANK_REPEAT; 3.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC ANK repeat.
SQ SEQUENCE 1224 AA; 136345 MW; 8BBE76BEC6B17E9C CRC64;

Query Match 7.8%; Score 117.5; DB 2; Length 1224;
Best Local Similarity 19.2%; Pred. No. 36;
Matches 65; Conservative 57; Mismatches 126; Indels 91; Gaps 10;

QY 2 TSIFAEQTVVEVVKSAIETADGALDLYNKYL-----DQV-----IPW 37

Db 876 TIIDLEQSLDNAQVSTSDADISRLNDELSTLKOQLSSKCLSEANSKIDLLKAELEE 935
QY 38 KTFD-----ETIKELSRFKQOEYSQASVLVGDIKVLMSQDKYFEATQTVYEWCGV 89
DB 936 KTIESASNPDLSSIPSLARSTSESTQNLGELSDELRELLDSERRRIKKLN-----NK 989
QY 90 VTQLLSAYILLPDEYNEKKAQAK-----DILIRILDGQVKLNE 129
DB 990 ITALMTEKNLNDQINAMKTSNAKNLPLPAAPVLEEPSLATGTDDVRISDLQALKDANE 1049
QY 130 ----AQKSLTSSQSFNNASGKLALDSQLTNDPSEKSSYFQSOVDRIKEAYAGAAAGI 185
DB 1050 RADAKERELTDALORIENLGAQISPLEAEUSS-IQDRLSATGEVIEKLAENQO----- 1102
QY 186 VAGPPGLIISIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVKQANKOID----- 237
DB 1103 -----LNQIAGAAGQSEBELITSLTRELEGLTENSQNLNAVKEDEKLRQIEDKVKELAI 1154
QY 238 ----AAKLKLAETI--AAIGEIKTETETTRFYVDYDML 270
DB 1155 ERQQAETLKLQRIIPDKIASYSDHVEKPRSVNESDDVM 1193

RESULT 22
HLY1_ECOLI STANDARD; PRT; 1023 AA.
ID HLY1_ECOLI
AC P09983;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemolysin, chromosomal.
DE Name=hlyA;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J96 / Serotype O4;
RC MEDLINE=85234404; PubMed=3891743;
RX Felmler T., Pellett S., Welch R.A.;
RT "Nucleotide sequence of an Escherichia coli chromosomal hemolysin.";
RL J. Bacteriol. 163:94-105 (1985).
RN [2]
RP SEQUENCE OF 1-44 FROM N.A.
RC STRAIN=2001;
RC MEDLINE=85258115; PubMed=3894051;
RX Nicoud J.-M., Mackman N., Gray L., Holland I.B.;
RA "Characterisation of HlyC and mechanism of activation and secretion of
RA haemolysin from E. coli 2001.";
RT FEBS Lett. 187:339-344 (1985).
RL
CC -1- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture by mechanisms not clearly
CC defined.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: The Gly-rich region is probably involved in binding
CC calcium, which is required for target cell-binding or cytolytic
CC activity.
CC -1- DOMAIN: The three transmembrane domains are believed to be
CC involved in pore formation by the cytotoxin.
CC -1- PTM: Palmitoylated by hlyC. The toxin only becomes active when
CC modified.
CC -1- MISCELLANEOUS: The hemolysin of E.coli is produced predominantly
CC by strains causing extraintestinal infections, such as those of
CC the urinary tract.
CC -1- SIMILARITY: Belongs to the RTX prokaryotic toxin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; M10133; CAA23975.1; -  
DR EMBL; X02768; CAA26546.1; -  
DR PIR; A24433; LEECA.  
DR InterPro; IPR001343; Hemlyan\_Ca\_bind.  
DR InterPro; IPR003995; RtxA.  
DR InterPro; IPR011049; Serralyen\_like\_C.  
DR Pfam; PF00353; HemolysinCabin; 6.  
DR Pfam; PF02382; RTX; 1.  
DR PRINTS; PR003113; CABNDNGRPT.  
DR PRINTS; PR01488; RTXTOXINA.  
DR PROSITE; PS00330; HEMOLYSIN CALCIUM; 4.  
KW Calcium; Cytolysis; Hemolysis; Lipoprotein; Palmitate; Repeat; Toxin;  
KW Transmembrane.  
FT TRANSMEM 237 259 Potential.  
FT TRANSMEM 267 326 Potential.  
FT TRANSMEM 364 410 Potential.  
FT DOMAIN 723 869 16 X REPEATS, GLY-RICH.  
FT REPEAT 723 728 1.  
FT REPEAT 732 737 2.  
FT REPEAT 741 746 3.  
FT REPEAT 750 755 4.  
FT REPEAT 759 764 5.  
FT REPEAT 768 773 6.  
FT REPEAT 777 782 7.  
FT REPEAT 786 791 8.  
FT REPEAT 795 800 9.  
FT REPEAT 806 812 10.  
FT REPEAT 816 821 11.  
FT REPEAT 825 830 12.  
FT REPEAT 834 839 13.  
FT REPEAT 843 848 14.  
FT REPEAT 855 860 15.  
FT REPEAT 864 869 16.  
FT LIPID 563 563 N(6)-palmitoyl lysine (By similarity).  
FT LIPID 689 689 N(6)-palmitoyl lysine (By similarity).  
FT VARIANT 6 A -> T (in strain 2001).  
SQ SEQUENCE 1023 AA; 109867 MW; 196D5C0A9A28B54D CRC64;

Query Match 7.7%; Score 116; DB 1; Length 1023;  
Best Local Similarity 20.5%; Pred. No. 36;  
Matches 69; Conservative 59; Mismatches 111; Indels 98; Gaps 11;  
QY 2 TSFAEQTVWVK-----SAIETADGALDLYNKYLQVPIPKTPTDIKELSRFKQE 53  
DB 155 TALSSMKIDELIKKQSGNVSSSELAKASIELINQLVDTA-----ASLNNVNSFSQ 207  
QY 54 YSQEASVL-----VGDYKVLMDSDQKYFEATQTVYEWGVVTVQLLSAYILL-PDEY 104  
DB 208 LNKLGSLVNTKHLGVGN-KLQNLPLNDIGAGLDTV---SGTLSAISAFILSNADAD 263  
QY 105 NEKKASAKDILIRLDD---GVKLNKAEQKSLTSSQSFNNASGKLALDLSQITNDFSE 161  
DB 264 TGTAAAGVELTKVLGNVGKISQYIIAQAAGLST----- 302  
QY 162 KSYFQSQVDRIRKEAYAGAAAGIVAGPFGGLIIS---YSIAAGVIEGKLIPELNNRKT 217  
DB 303 -----AAAGLIASVTVTLAISPLSLFIADKFKRANKIEYSQFFK 344  
QY 218 VQNFTSLSTVQKANKIDAAKLKATEIAAIGIKETETTRFYVDYDMLMLSLKGA 277  
DB 345 LGYDGDLSLAFFHKTGAIDASLTRISTVLASVSSGISAAATSLV---GAPVSAIVGA 400  
QY 278 -----AKMINTCNEYQQRHGK 294  
DB 401 VTGIISGILEASKOAMPEHVASKWADVIAEWKKGK 437

RESULT 23  
Q7XEH4  
ID Q7XEH4 PRELIMINARY; PRT; 2033 AA.

AC Q7XEH4;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Putative centromere protein.  
GN ORFNames=OSJNBa0075H14.2;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA The Rice Chromosome 10 Sequencing Consortium;  
RT "in-depth view of structure, activity, and evolution of rice chromosome 10."  
RL Science 300:1566-1569 (2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;  
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB017095; AAP53815.1; -  
DR Gramene; Q7XEH4; -  
DR InterPro; IPR001917; Aminotrans II.  
DR PROSITE; PS00599; AA\_TRANSFER CLASS 2; UNKNOWN 1.  
SQ SEQUENCE 2033 AA; 231439 MW; ACFD716945B5C462 CRC64;  
Query Match 7.7%; Score 116; DB 2; Length 2033;  
Best Local Similarity 18.7%; Pred. No. 82;  
Matches 67; Conservative 62; Mismatches 106; Indels 124; Gaps 11;  
QY 7 EQTVVVKSAIE-----TADGA-----LDLYNKYLQVPIPKTPTDIKELSRFKQESQEA 58  
DB 477 EQKVQMLQDLEQKQEQADSAHAQLQDCENRH-----TQTEADLHFRKHLHSQ-- 524  
QY 59 SVLVGDIKVLMDSDQKYFEATQTVYEWGVTVQLLSAYILLPDEYNEKKA-----SAQK 113  
DB 525 -----LEEVIKLTENLDRSTKSELENAKLDLENTRELKSTLDLNLSEK 571  
QY 114 DILIRLDDGVKKLNKAEQKSLTSSQSFNNASGKLALDLSQLT-----NDFS 160  
DB 572 DAVLQQQSLAKISELEQLSKTQLELNKSEKQMLLELEITQKSEMSDSLTLKDET 631  
QY 161 EK-----SSYFQSQ--VDRIRKEAYAGAAAGIVAGPFGGLIISYIAAGVIEGK 206  
DB 632 EKVQAEATSLMSMESMYSQSEEVNRLHLE----- 661  
QY 207 LIPELNNRKTQVNFSTLSATVQKANKIDAAKLKATEIAAIGIKETETTRFYV-- 264  
DB 662 -IEKLNFKLNELENLSSELNLTILLNAEKDADTLKRNQQLSVIRISLESLSKLQALEK 720  
QY 265 -----DYDDLMLSLKGAKK-----MINTCNEYQQRHGKKT 297  
DB 721 IEKVQMLEQELKHKEVDLSLQISIQDEAHKSEGEAALLANTNLNSESQEEVNRLLT 779  
RESULT 24  
SR68 CAEEL  
ID SR68 CAEEL STANDARD; PRT; 622 AA.  
AC Q208Z2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Probable signal recognition particle 68 kDa protein (SRP68).  
GN ORFNames=F55C5.8;  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Harris B.;

RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Signal-recognition particle assembly has a crucial role  
CC in targeting secretory proteins to the rough endoplasmic reticulum  
CC membrane. SRP68 binds the 7S RNA, SRP72 binds to this complex  
CC subsequently. This ribonucleoprotein complex might interact  
CC directly with the docking protein in the ER membrane and possibly  
CC participate in the elongation arrest function (by similarity).  
CC -1- SUBUNIT: Signal recognition particle consists of a 7S RNA molecule  
CC of 300 nucleotides and six protein subunits: SRP72, SRP68, SRP54,  
CC SRP19, SRP14 and SRP9 (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: Belongs to the SRP68 family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL: Z78198; CAB01573.1; -  
CC PIR: T22716; T22716  
CC wormPep: F55G5.8; C820875.  
CC InterPro: IPR008941; TPR-like.  
CC KW Hypothetical protein; Ribonucleoprotein; RNA-binding;  
CC KW Signal recognition particle.  
CC SQ SEQUENCE 622 AA; 70574 MW; A7B8808E46169636 CRC64;  
CC  
CC Query Match 7.6%; Score 115.5; DB 1; Length 622;  
CC Best Local Similarity 21.2%; Pred. No. 22;  
CC Matches 68; Conservative 56; Mismatches 109; Indels 87; Gaps 15;  
CC  
CC QY 41 DETIKELSRFKQEQYSQASVLVGDIKVLLMSQDKYFEATQTVYEWGCVTQLLSAY--- 97  
CC Db 251 DKLISEM-RASATSAEVTWIEGGAQSTVDDE-----KAKQVQEWKQTEVELAQCTPTK 304  
CC QY 98 --ILLPDEYNEKASAKQDKILRIIDDDGVKKLNE-AQKSLTSSQSF-----NNASGKL 148  
CC Db 305 EKMALEP-----EKATADTRDAIDRIISDIRKSSSENADTVLQSIKAYLEFLKMGNTASRY 360  
CC QY 149 LALDSQLTNDFSEKSYFOS-----QVDRIRKEAYAG 180  
CC Db 361 LAI---IDNTKSEKSKPQDLRLLYDSVIEIYKEVAEIPGADHDKNLIQAPEVKVEYRA 417  
CC QY 181 AAGIVAGPFLGLIISYIAAGVI-----EGKLIPELNNRLKTVQNFETLSATVK 230  
CC Db 418 FRCFYMASSYSALHKYSEAAALFDRTVSRVQDAEGL-----KKLKSSSFITNETQSSIN 472  
CC QY 231 QANKDIDAAK--LKLATEIAAIGEIKTETETTRFFVYDDMLSLKGAAKKMINTCNEY 288  
CC Db 473 ELRSEVESAKVTVAARLASAGDVKTDSLAKI-ID-----KRLLETVTNEW 519  
CC QY 289 QQ---RHGKKTLEFVPDVAS 305  
CC Db 520 QWDVRSNLKDKKTIIP-VAS 538  
CC  
CC RESULT 25  
CC Q97WH8 PRELIMINARY; PRT; 587 AA.  
CC AC Q97WH8;  
CC DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
CC DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
CC DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
CC DE BPS2 protein homolog (Bps2).  
CC GN Name=Bps2; OrderedLocusNames=SSO2241;  
CC OS Sulfolobus solfataricus.  
CC OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
CC CC Sulfolobus.  
CC OX NCBI\_TaxID=2287;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
RX MEDLINE=21332296; PubMed=11427726;  
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
RA Aweze M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
RA De Moors A., Brauso G., Fletcher C., Gordon P.M.K.,  
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;  
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
DR EMBL: AE006828; AAK42408.1; -  
DR PIR: A90394; A90394.  
KW Complete proteome.  
SQ SEQUENCE 587 AA; 68436 MW; B90DBC1E19C05E86 CRC64;  
CC  
CC Query Match 7.6%; Score 115; DB 2; Length 587;  
CC Best Local Similarity 20.3%; Pred. No. 22;  
CC Matches 61; Conservative 66; Mismatches 103; Indels 70; Gaps 14;  
CC  
CC QY 10 VEVVKSALETADGALDLYNKYLDQVIPWKTFDETI--KELSRFKQEQYSQASVLVGDIKV 67  
CC Db 42 ISLTSSIK---AEDLLNVEADSGYVEALDNKLYRRIRKIRNGELGEKNLMDDDRA 97  
CC QY 68 LLMSQDKYFEATQTVYEWGCVTQLLSA-----YILLPEYNEKASAKQDKILRIILD 121  
CC Db 98 LLL-----TYFSPENR-----LVTOILSGDGNVWFISTTSKINEIK--AKKEELQKLT 145  
CC QY 122 DGVKLNBAQKSLTSSQSFNN---ASGKLALDSQLTNDFSEKSYFOSQVDRIRKEAY 178  
CC Db 146 AEINARDELQK-----KYNIREIQAIRAID-----EIDKLEKE-- 181  
CC QY 179 AGAAAGIVAGPFLGLIISYIAAGVIEGKLIPELNNRLKTVQNFETLSATVKQAKKIDA 238  
CC Db 182 RESSSNIVAK-----TTVTIT--LFRONKINEILNKIKVKKDELANLEFALKIEEELQN 234  
CC QY 239 AKLKATEIAAIGEIKTETETTRFFVYDD-----LMLSLKGAAKKM-----INTCN 286  
CC Db 235 KESKVSFDIKTQLEKEMEINEKCLKTNDRSELSIELKLVLEEVNEDRHHLDTCN 294  
CC  
CC RESULT 26  
CC O64067 PRELIMINARY; PRT; 478 AA.  
CC AC O64067;  
CC DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
CC DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
CC DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
CC DE Hypothetical protein yond.  
CC GN Name=yond;  
CC OS Bacteriophage SPBc2.  
CC OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
CC OX NCBI\_TaxID=66797;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Mavel C.,  
CC RA Karamata D.;  
CC RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
CC DR EMBL: AF020713; AAC13027.1; -  
CC DR PIR: T12818; T12818.  
CC KW Hypothetical protein.  
SQ SEQUENCE 478 AA; 55050 MW; 6FF7495A957D4A4F CRC64;  
CC  
CC Query Match 7.6%; Score 114.5; DB 2; Length 478;  
CC Best Local Similarity 22.6%; Pred. No. 19;  
CC Matches 57; Conservative 41; Mismatches 73; Indels 81; Gaps 12;  
CC  
CC QY 25 DLYNKY-IDQVIPWKTFDETIKELSRFKQEQYSQASVLVGDIKVLLMSQDKYFEATQTV 83  
CC Db 268 DVIDTYFTVNVYSWS--DENSVD-KYFKFNTR-----TGDVTSIDFDSKTEVF---MT 315  
CC QY 84 YEWGCVTQLLSAYILLPDEYNEKASAKQDKILRIIDDDGVKLNBAQKSLTSSQSFNN 143  
CC  
CC RP SEQUENCE FROM N.A.









```

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000631; EAA21743.1; --
DR GO; GO:0005525; F:GTP binding; IEA.
DR InterPro; IPR005289; GTP-binding.
DR InterPro; IPR002917; MMR_HSR1.
DR Pfam; PF01926; MMR_HSR1; 1.
DR TIGRFAMs; TIGR00650; MG442; 1.
SQ SEQUENCE 779 AA; 92336 MW; 7F51893C396B0B52 CRC64;

Query Match 7.6%; Score 114.5; DB 2; Length 779;
Best Local Similarity 20.5%; Pred. No. 33;
Matches 71; Conservative 51; Mismatches 128; Indels 97; Gaps 14;

QY 24 LDLYNKYLDQVTPKMTFDTIKELSRFKQYSEASVVGDIKVLMDSDQKYFEATQV 83
DB 92 LGKYGVETVFQTFDEYFN--SRKEDYEQ--LEAGLKLI-LTENTNTYKDBIRKI 145

QY 84 YEWCGVVTQLLSAYILLFDEYNEKASAKQD-----ILIRLDD----- 122
DB 146 ME-----KSY-----KSEYELKKEEEDSKISHDVHENLKVTFVKKNIIDNFLTQY 193

QY 123 --GVKLINEAQKSLTSSQSNFNASGKLALDLSLTND---FSEKSYF-----OSQVD 171
DB 194 STYINKLKESEILSKLVNFKNGSTTSNIEENNDQVCLDEKSEKYSNDEKSKLR 253

QY 172 RIRKEAYAGAAGIVAGPFGLIISYISIAAGVIEGKLPDLN-----NRLKTVQNF 223
DB 254 EINKPGDAENA-----LEFYKGSNKENKLIBEVNFFPKQGYAEKRISTFSD 302

QY 224 SLSATVQKQNK-----DIDAKLKLATEIAAIGIKTETTRFYVDYDMLSL 274
DB 303 QMFANVKIENKLLIKGEYDNDGGLKLEKGEVINSYNNDEGTEFLKNNVSVLELF 362

QY 275 KGAACKMI-----NTCNE-----YQORHGKKTLEFVDPD 302
DB 363 KENVKASLYKKENIIKNNLENLLCGRVKVVHWFPPKMKRIIMKIPD 409

RESULT 33
Q869B8 PRELIMINARY; PRT; 1885 AA.
AC Q869B8
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Kinesin-related protein K4.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98359834; PubMed=9693369;
RA de Hostos E.L., McCaffrey G., Sugang R., Pierce D.W., Vale R.D.;
RT "A developmentally regulated kinesin-related motor proteins from
RT Dictyostelium discoideum."
RL Mol. Biol. Cell 9:2093-2106(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Suyama E., Sutoh K.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB102780; BAC56912.1; --
DR HSP; P33173; 1161.
DR DictyBase; DB0191404; ksmD.
DR GO; GO:0005875; C:microtubule associated complex; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PRO0380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN2; 1.

```

```

KW ATP-binding; Microtubule; Motor protein.
SQ SEQUENCE 1885 AA; 218215 MW; 15686AED3B007EC7 CRC64;

Query Match 7.6%; Score 114.5; DB 2; Length 1885;
Best Local Similarity 20.7%; Pred. No. 95;
Matches 76; Conservative 51; Mismatches 109; Indels 132; Gaps 14;

QY 40 FDETIKELSRFKQYSEASVVGDIKVLMDSDQKYFEA 79
DB 942 FDSQDQNSQIKSYNQLESTLTLAQSENQRLTENKQFITSLNEIKSLFNSTQQOK-ET 1000

QY 80 TQTVYEWGVVQ---LLSAYILLFDEYNEKASAKQDILIRLDDGVKLINEAQKSLLT 136
DB 1001 IOLFEFOEKENQFDSLLTNYNQLFKYNDLAYSNESRL--EFDQFKDSDNOSIOSLES 1058

QY 137 SSQFNNASGKLL---ALDSQLTN-----DFSSEKSYFQSQVD 171
DB 1059 LERSKSENDNLLQSSLLKSLQLESIEKQKQDQLPIQLESKKCELSKLSQFSQSTK 1118

QY 172 RIRKEAYAGAAGIVAGPFGLIIS--YSTAAGVIEGKLI---PELNN----- 213
DB 1119 QVTQ-----LLISVDQYKISTNKLSEQISDRNEEINNLKLKAEINALK 1162

QY 214 -----RLKTVQNFFTSLSATVQKQNK---IDAAK 240
DB 1163 EENISLKDQTLKLVKAPKQSDTRDKMKLEKLEKFNDAKLAQAIQDKTQIOSEK 1222

QY 241 LKLATEIAAIGIKTETTRFYVDYDMLSLKLG-----AAKQNTCNEYQORH 292
DB 1223 QSLEREIKDLKRSHTSTET-----ELDKLKLTHLAADVKSDFIALNKSEVILTKSOQL 1277

QY 293 GKKTLEFV 300
DB 1278 -KSTIIEL 1284

RESULT 34
Q9H6Q7 PRELIMINARY; PRT; 720 AA.
AC Q9H6Q7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ21979.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025632; BAB15196.1; --
FT NON_TER 720
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Query Match 7.5%; Score 114; DB 2; Length 720;
Best Local Similarity 20.7%; Pred. No. 33;
Matches 62; Conservative 56; Mismatches 111; Indels 70; Gaps 10;

QY 26 LYNKYLDQVTPKMTFDTIKELSRFKQYSEASVVGDIKVLMDSDQKYFEATQ 81
DB 457 LYKEYLSEKTEKTEKTEKTEKTEKTEKTEKTEKTEKTEKTEKTEKTEKTEKTEK 512

QY 82 TVYEWGVVVTQLLSAYILLFDEYNEKASAKQDILIRLDDGVKLINEAQKSLTSSQSF 141
DB 513 ILAENSRRITVL-----QVNEKSLIRQYTTLVLE-LERQLRKENEKOKN----- 554

QY 142 NNASGKLALDLSLTNDPSEKSYFQSQVDRIKKEAYAGAAGIVAGPFGLIISYIAAG 201
DB 555 -----ELLSMEAEV-----CEKIGCLQ-----RFKEMAIKFAA-----LQKVDNVSLS 595

```



```
QY 202 VIE-----GKLIPELNRLKTVQNFPTSLSATVQKANKDIDAALK 242
Db 596 EELANKQYNELTAKYRDILQKNMLVQRTSNLEHLECEINSLKEQVESINKELEITKEK 655
QY 243 LAT-----EIAAIGEIKETETTRFRFVVDVDDMLSLKGAACKMINTCNEYOORHGCK 295
Db 656 LHTEQAEQETKLGNESSMDKAKSITNSDIVISIKKKKKKKKKKKKKKKKKKKKKKK 714

RESULT 35
Q8CEP10
ID O8CEP10 PRELIMINARY; PRT; 1189 AA.
AC O8CEP10
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chromosome segregation SMC protein.
GN OrderedLocuNames=SE0909;
OS Staphylococcus epidermidis
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
  Qian Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
  Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
  Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
DR EMBL; AE016747; AA004506.1; -.
DR HSP; P58301; 1P2T.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0007059; F:GTP binding; IEA.
DR GO; GO:0007059; P:chromosome segregation; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR005289; GTP-binding.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR010935; SMC hinge.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF06470; SMC_hinge; 1.
DR Pfam; PF02463; SMC_N; 1.
DR TIGRFAMs; TIGR00650; MG442; 1.
DR Complete proteome.
SQ SEQUENCE 1189 AA; 137542 MW; 0F6F95636BE31BC8 CRC64;

Query Match 7.5%; Score 114; DB 2; Length 1189;
Best Local Similarity 18.6%; Pred. No. 59;
Matches 62; Conservative 59; Mismatches 130; Indels 82; Gaps 9;

QY 7 EQTVVVK-----SAIETADGALDLYNKYLDQVPWKTFDETIKLSRPFQKQYASV 60
Db 163 EESAGVLKYKRAESIQKLDHTEDNLNRVEDILYDLGRVPLKEAATAKEYKQLSKE 222

QY 61 LVGDIKVLMDSQKYEATQTVVWC-----GVVTQLLSAYILLFDEYNE 106
Db 223 MEQSDVITVTSIDHYTQDNRQLRDLRNLHLSQQAQEGQQAQINQLLQKY----- 273

QY 107 KKASAKDILIRILDGCVKKNLAKQSLLTSSQSFNNASGKLLALDSQLTND-----FS 160
Db 274 -KGRQON-----DYDIEKLN-----YELVKATNEYQLSGKLNVLBERKKNQSETNARYE 323

QY 161 EKSSYFQSDVRIRKEAYAGAAAGIVAGPFGLIISYSIAAGVTEGKLIPELNRL-----K 216
Db 324 EELDNLESQSDISIKNEKAQNE-----KLLAELKNKQKQLNK 359
```

```
QY 217 TVQNFTSLSATVQKANKDIDAALKLATEIAAIGEIKETETTRFRFVVDVDDMLSLK 276
Db 360 EVOELESLLYISDEQDEKLEIKNSYLLMSQSDVNNDIRFLERTINENAKSKRLDS 419
QY 277 AAKKMINTCNEYQO-----RHGKKTLFEV 300
Db 420 RLVEAFNQLKDIOQNTQTQKEYQSSKSKSEKV 452

RESULT 36
Q7RD43
ID Q7RD43 PRELIMINARY; PRT; 1611 AA.
AC Q7RD43
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Names=PY05582;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteza M.,
  Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
  Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
  Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
  Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
  Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
  Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
  van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
  Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
  Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
  parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
  EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
  preliminary data.
CC EMBL; AABL01001787; EAA17620.1; -.
DR Hypothetical protein.
SQ SEQUENCE 1611 AA; 188293 MW; CD0493CC4A9BC5D1 CRC64;

Query Match 7.5%; Score 114; DB 2; Length 1611;
Best Local Similarity 21.7%; Pred. No. 85;
Matches 70; Conservative 60; Mismatches 111; Indels 82; Gaps 15;

QY 24 LDLYNKYLDQVPWKTFDETIK-ELSRFQKQY-----SQEASVLVGDIKVLML----- 70
Db 799 LNIIRNLYLMKK-----NEKIKFENKLLKDYDNLKSKYKNVLPKLCIKFLLKYYTKD 852

QY 71 -DSQDKYFEATQTVVWCVVVTTQLLSAYILLFDEYNEKASAKDILIR-ILDDGVKKLN 128
Db 853 RSYDKKYE-----NMRHLQSEKKNLMLIEKKYQEQNNININMMLSDLKQLQ 903

QY 129 EAQKSLTSSQSFNNASGKLLALDSQLTNDPSE-KSSYFQSQVD-----RIRKE 176
Db 904 EENESIISNNNSYKTEVETINSKYQLQNDFNKIKSEHEKLIKIEHKIKRENENIKIEKE 963

QY 177 AYAGAAGIVAGPFGLIISYSIAAGVIEG-----KLIPELNRLKTVQNFPTSLSATVK 230
Db 964 TLIKELGDTKAKYFNM-----TGILQGBEKMYAKKIKLEELKEKEKEKKNIIINDVK 1016

QY 231 QANKDIDA-AKL-----KLATEIAAIGEIKETETTRFYVD-----YDDLML 271
Db 1017 D-----EIDFTFKILEKENENHRIKELKELUNKEEKYENTQINFDSLKKEFEKSFDEVQI 1073

QY 272 SLLKGAACKMINTCNEYQORHGK 294
Db 1074 IL-----KEMIQKEYTNSYNK 1091
```

Db 344 IDRLSGKLNTRTRVFLAEAS 366  
: ||: : ||

## RESULT 37

Q9Z429 PRELIMINARY; PRT; 538 AA.  
ID Q9Z429  
AC Q9Z429  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Meth1-accepting chemotaxis protein.  
GN Name:ahy;  
OS Pseudomonas putida.  
OS Plasmid NAH7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G7;  
RX MEDLINE=99255564; PubMed=10322041;  
RA Grimm A.C., Harwood C.S.;  
RT "Nahy, a catabolic plasmid-encoded receptor required for chemotaxis of  
RT Pseudomonas putida to the aromatic hydrocarbon naphthalene.";  
RL J. Bacteriol. 181:3310-3316(1999).  
DR EMBL; AF100302; AAD13223.1; --  
DR HSSP; P02942; 1QUT.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0004871; P:signal transducer activity; IEA.  
DR GO; GO:0006935; P:chemotaxis; IEA.  
DR GO; GO:0007165; P:signal transduction; IEA.  
DR InterPro; IPR004089; Chmtaxis\_trans.  
DR InterPro; IPR003660; HAM.  
DR InterPro; IPR004090; Me\_chemotaxis.  
DR InterPro; IPR000727; T\_SNARE.  
DR Pfam; PF00672; HAM; 1.  
DR Pfam; PF00015; MCFsignal; 1.  
DR PRINTS; PR00260; CHEMTRNSDUCR.  
DR SMART; SM00283; MA; 1.  
DR SMART; SM00111; CHEMOTAXIS\_TRANSUC\_2; 1.  
DR PROSITE; PS00885; HAM; 1.  
DR PROSITE; PS00192; T\_SNARE; 1.  
DR Plasmid.  
KW Complete proteome.  
SQ SEQUENCE 538 AA; 58716 MW; C91589BE65F45E56 CRC64;

Query Match 7.5%; Score 113.5; DB 2; Length 538;  
Best Local Similarity 21.9%; Pred. No. 25;  
Matches 84; Conservative 56; Mismatches 118; Indels 125; Gaps 20;  
QY 9 TVEVVK-SAIETADGALDLYNKYLDQVTPWKTFDETIKLSRFKQYEQEASVLGDIKV 67  
DB 23 TIELMGFSALQR--GVASLNTVYLDRVF-----LRDLKTADLYA-----V 62  
QY 68 LMDSQDK-----YFEATQTVVWGVVTTLLSAYILLFDEYNEKKSAAQK---DILI 117  
DB 63 KIVDSHKSARGMTVAQAEQVKGQIDMLWHAY-----QTKKIDEEQSRVDALA 116  
QY 118 RILDDGVKLINEAQKSLTSSQSFNNASGKLALDSQLTND-----PSEKSSYF-QSQ 169  
DB 117 KLVDQAQDPI-ERLKGIL-----ERGDKAALDTFVENMYPLDPLSEGLSHLTQIQ 167  
QY 170 VDIRKQAYAGAA-----GIV--AGPFGLLIYSI-----AAGVI 203  
DB 168 VEE-SKRAYDAAVVLYDSRTMLALLLLGLICGGVFATRLRSIHPITLTKDAARVA 226  
QY 204 EGKLIPEL-----NNRLKTVQNFSTLSATVKQANKDIDAALKATEIAAIGETETET 259  
DB 227 LGDLSQSIQVSGRNEVDVQSQVQAMQANRLNTLQDIQGSAAQLA---AAAELOQTATES 283  
QY 260 TR--FYVDYDDLMS-----LLKGAARKKQWINTCNE 287  
DB 284 TAQGIHQRQNDQMQAATAVTMSAAVDEADNANRTSNASHEAMDLDAGGRKQVMLTRET 343  
QY 288 YQORHGK-----KTLFEPDVA 305

## RESULT 38

Q8ZON2 PRELIMINARY; PRT; 727 AA.  
ID Q8ZON2  
AC Q8ZON2  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE All0059 protein.  
GN OrderedLocusNames=all0059;  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003581; BAB77583.1; --  
DR PIR; AC1814; AC1814.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.  
DR InterPro; IPR003856; LPS\_Wzz\_MPA.  
DR Pfam; PF02706; Wzz; 1.  
KW Complete proteome.  
SQ SEQUENCE 727 AA; 82555 MW; B93994D1F4656D8 CRC64;

Query Match 7.5%; Score 113.5; DB 2; Length 727;  
Best Local Similarity 22.2%; Pred. No. 36;  
Matches 75; Conservative 58; Mismatches 134; Indels 71; Gaps 15;  
QY 5 FAEQTVVVKSAIETADGALDLYNKYLDQVTPWKTFDETIKLSRFKQYEQEASVLGVD 64  
DB 194 FINARLFVTKQQRARLEQFRKKNLLDPELOSQVLIKSLTK-TQEQLQLTQLQD 252  
QY 65 IKVLLMDSQDKYFEATQTVVWGVVTTLLSAYILLFDEYNEKKSAAQKDIILRILD 121  
DB 253 INSRYNLEERITEASQK-----ALVSMELAQSSRYKTLTSELQKTEQSLAKEQL-RYTD 306  
QY 122 DG--VKKLINEAQKSLT-SSQSFNNASG-----KLALDSQLTNDF----- 159  
DB 307 DSPIVQSLKQRRSQTLLVRQELKRLTAEKLTQTTPEPQOLGVDPNLVEEFVQVQTAL 366  
QY 160 ----SEKSSYFQSQVDRIKEAYAGAAAGIVAGPFGLLIYSIAGVIEGKLIPELNRL 215  
DB 367 GLIANEKS--LRSEQRIRSE-----LSKYPSLIAEQ-----RLPEVETQR 407  
QY 216 KTVQNF--TSLSTVTKQANKDIDAALKATEIAAIGETETETTRFYVDYDDL----- 269  
DB 408 KTLQMLQAQSQMGKMLAHGGFD---WLVEEANLG---TYVGNDRFLVFGVMTGPI 460  
QY 270 ---MLSLKGAARKKQWINTCNEYQQRHGKTLFEPDVA 304  
DB 461 LGVVALTWGMRHRIIHSADQLQKVSRLRLGTVPKLA 498

## RESULT 39

Q9YVT6 PRELIMINARY; PRT; 1127 AA.  
ID Q9YVT6  
AC Q9YVT6  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Hypothetical protein MSV156.



```

QY 193 IISYSIAAGVIEGKLIPELNNRLKTVQNFSTL-----SATVQKANKD 235
Db 436 -IS-NLSIDFQETKL-KENGDKLSTLKELLESKNOKNIERDKKELDEVNSKIBQIERD 492
QY 236 IDAAK-----IKLATEAAGEI-KTETETTRFYVDYDMLSLKGAACKMINTCNEYQQ 290
Db 493 VNOSKNGYEIGVEKINEAEANKRIESTK-----ELIQTIQNLSTQBELLK 542
QY 291 --RHGKK 295
Db 543 NIEGKNK 549
RESULT 42
ID Q6C359 PRELIMINARY; PRT; 1906 AA.
AC Q6C359;
DT 01-OCT-2004 (TREMREL. 28, Created)
DT 01-OCT-2004 (TREMREL. 28, Last sequence update)
DT 01-OCT-2004 (TREMREL. 28, Last annotation update)
DE Similar to DEHAOC096589 Debaromyces hansenii IPF 1836.1.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG GENOLEVURES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisarame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RN Nature 430:35-44(2004).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382132; CAG77705.1; -.
SQ SEQUENCE 1906 AA; 210346 MW; F6ED7D1AF7B7562B CRC64;

Query Match 7.5%; Score 113.5; DB 2; Length 1906;
Best Local Similarity 22.1%; Pred. No. 1.1e+02;
Matches 66; Conservative 48; Mismatches 116; Indels 69; Gaps 12;

QY 7 EQIVVVVKSIAETADGALDLYNKYL----DOVIPWTKTFDTIKELSRFRQYTSQASVLV 62
Db 1195 EKELETKTSLETKTAELTKSKELTAKSDEA---TTSYAKVKELTSSAALEKKQTLLK 1251
QY 63 GDIKVLMDSDQKYFEATQTVYEMCVGVVTVQLLSAYILLFDEYNEKKAQAKQILYILD 122
Db 1252 AMADNLTKDLAEK-----TKELVAKSEL--ESSNTSKEEVDVLTKKLSD 1295
QY 123 GVKKLNEAQSLLTSSQFNNSKLLALDSQLTNDFSEKSSVFQSOVDRIKAEVAGAA 182
Db 1296 ATAFAVELKKS---SQAAETASSKVSALEAKUTK-ASESS---KAEIDKVNK----- 1341
QY 183 AGIVAGPPLIISYSIAAGVIE-----GKLIPELNNRLKTVQNF---FTLSATVQKA 232
Db 1342 -----LLSSPFKEKLTQSKDHDHSTEVSKLTQEVRESTLKAENFEHDISSLLKDLQA 1392

```

```

QY 233 NKDIDAARKLATEIAIGSIKTTETETTRFYVDYDMLSLKGAACKMINTCNEYQOR 291
Db 1393 EKEDARLTETDTSIK-----EMENERTSLTKDAD-----SATKELTNKVSMLQTK 1438
RESULT 43
ID Q45614 PRELIMINARY; PRT; 3102 AA.
AC Q45614; P91824; Q9TZR4; 06, Created)
DT 01-JUN-1998 (TREMREL. 22, Last sequence update)
DT 01-OCT-2002 (TREMREL. 22, Last annotation update)
DT 01-OCT-2004 (TREMREL. 28, Last annotation update)
DE C. elegans LAM-3 protein (Corresponding sequence T22A3.8) (Laminin
DE alpha chain) (Hypothetical protein T22A3.8).
OS Name=lam-3; Synonyms=lama1/2; ORFNames=T22A3.8;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Telodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RA "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Barlow K.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=22668894; PubMed=12783803;
RA Huang C.C., Hall D.H., Hedgecock E.M., Kao G., Karantz V.,
RA Vogel B.S., Hutter H., Chisholm A.D., Yurchenco P.D., Wadsworth W.G.;
RA "Laminin alpha subunits and their role in C. elegans development.";
RL Development 130:3343-3358(2003).
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurray A.A.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL008585; CAA15432.3; -.
DR EMBL; Z81125; CAA15432.3; JOINED.
DR EMBL; AL008585; CAB03385.3; JOINED.
DR EMBL; Z81125; CAB03385.3; -.
DR EMBL; AF074902; AAC26793.1; -.
DR PIR; F87908; F87908.
DR PIR; T23064; T23064.
DR PIR; T43291; T43291.
DR HSSP; P02468; INPE.
DR WormPep; T22A3.8; CE31067.
DR GO; GO:0005578; C:extracellular matrix; IEA.
DR GO; GO:0005606; C:laminin-1; IEA.
DR GO; GO:0005102; F:receptor binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030155; P:regulation of cell adhesion; IEA.
DR GO; GO:0030334; P:regulation of cell migration; IEA.
DR GO; GO:0045995; P:regulation of embryonic development; IEA.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR008979; Gal_bind like.
DR InterPro; IPR009030; Grow_fac_recept.
DR InterPro; IPR000034; Laminin_E.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR009254; Laminin_I.
DR InterPro; IPR010307; Laminin_II.
DR InterPro; IPR008211; Laminin_N.

```

```

DR InterPro: IPR003129; TSP_N.
DR Pfam: PF00052; Laminin_B; 2.
DR Pfam: PF00053; Laminin_EGF; 17.
DR Pfam: PF02210; Laminin_G; 3.
DR Pfam: PF06008; Laminin_I; 1.
DR Pfam: PF06009; Laminin_II; 1.
DR Pfam: PF06009; Laminin_II; 1.
DR Pfam: PF00055; Laminin_N; 1.
DR PRINTS: PR00011; EGF_LAMININ.
DR ProDom: PD003031; Laminin_B; 2.
DR SMART: SM00180; EGF_Lam; 17.
DR SMART: SM00281; LamB; 2.
DR SMART: SM00282; LamG; 4.
DR SMART: SM00136; LamNT; 1.
DR PROSITE: PS00022; EGF_1; 12.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 14.
DR PROSITE: PS50025; LAM_G_DOMAIN; 3.
DR PROSITE: PS50025; LAM_G_DOMAIN; 3.
DR Hypothetical protein; Laminin EGF-like domain.
KW SEQUENCE 3102 AA; 343790 MW; 2EE9305D5C591FAB CRC64;

Query Match 7.5%; Score 113.5; DB 2; Length 3102;
Best Local Similarity 20.8%; Pred. No. 2e+02;
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;

QY 25 DLYNKYLDQVLPW-KTFDETIKELSRFKQYEQASVLVG-----DIKVLIMDS 72
DB 1881 DLKNR-IDVLEQMNDYRETIYDVKKDTADAEKMSLVGVGRINRYKEVSEIEKURVEA 1939

QY 73 QDK-----YFEATQTV-----YEWCGVVTQLLSAYILLFDEY 104
DB 1940 EDQIAYSNSIEKARSEELNMWFEDEKINNTWLAELPDVLEQCNITLL---YSQILDEY 1996

QY 105 NEK-----KASAQKIDILIRLD-----DGVKYLNEAOKS 133
DB 1997 DEEYVQTAGRAHAEKLEVQAOK-IVDRFVDTRTETENPLKASHAYENIVEALKKNATEAVDS 2055

QY 134 LRTSSQSFNNASGKLALDSQLTWDFSEKSYFQSOVDRIKKEAYAGAAAGIVAGPGLI 193
DB 2056 AAEEA-----AVSKMLGSESGDANES--LRSQLKLNESLSN----- 2097

QY 194 ISYSIAAGVIE--GKLLPELNNRLKTVQNFPTSLSATVQKQKIDDAKLLATEIAAIG 251
DB 2098 VDSNNAVKIVBELKKEKKDLTDLGLHNLKTSI-----VKRLG 2136

QY 252 EIKTETETTFPYDYLMLSLKGAAKKM-INTCNEYQQRHGKKT 296
DB 2137 VIKNEASS--WDDKHDRMHSILKNGAKTAHRSANVKESEGIKT 2179

RESULT 44
Q73AJ3 PRELIMINARY; PRT; 465 AA.
AC Q73AJ3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE Flagellin.
GN OrderedLocusNames=BCE1780;
OS Bacillus cereus (strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=222523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714;
RA Raiko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L., Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F., Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic adaptations and a large plasmid related to Bacillus anthracis px01.";
RL Nucleic Acids Res. 32:977-988(2004).
DR EMBL; AE017269; AAS40708.1; --
DR TIGR; BCE1780; --
SQ SEQUENCE 465 AA; 49855 MW; E4C307E7717381D5 CRC64;

Query Match 7.5%; Score 113; DB 2; Length 465;
Best Local Similarity 22.0%; Pred. No. 23;
Matches 72; Conservative 53; Mismatches 116; Indels 86; Gaps 15;

QY 6 AEOTVEVKSIAETADGALDLYNKYLDQV---IPWKTFTDETIKELSRFK-----Q 52
DB 184 AKATI-TAKDAFDAKADSDAAKKEIDAAAKOTPSKNDAAQLAKEYIEAKATLATLKPTDA 242

QY 53 EYSQEASVL-----VGDIKVLIMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNE 106
DB 243 TYAAKAAELDAATTAALNDNAKVLVDGYEKKLTTTKTEA-----BYTA 285

QY 107 KKAQAQKIDILIRLDGKVKLNEAOKSLTSSQSF-NNASGKLLA---LSQLT-----N 157
DB 286 AKEQSTK-----STAAADLVTKYETAKSNALGNDIAKEYLEAKTAYEANKN 331

QY 158 DFSEKSYFQSOVDRIKKEAYAGAAAGIVAGPGLIISYSIA--AGVIECKL--IPELNN 213
DB 332 DISSKR-FAEATELNKIDITANKAAKV-----LVETYEKAKTAGTTEKSLVAVDKIDE 384

QY 214 RLKTVQNFPTSLSATVQKQKIDDAKLLATEIAAIGEI-----KTETETTFPYDYL 267
DB 385 ALKTIADNRATLGATLNRDLDFVNNLKSQASASASQIEDADMAKEMSEMTKF----- 439

QY 268 DMLSLKGAAKKMINTCNEYQQRHGK 294
DB 440 ----KILNEAGISMLSQANQTPQWVK 462

RESULT 45
AAS40708 PRELIMINARY; PRT; 465 AA.
ID AAS40708
AC AAS40708;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Flagellin.
GN BCE1780.
OS Bacillus cereus (strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=222523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714;
RA Raiko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L., Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F., Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic adaptations and a large plasmid related to Bacillus anthracis px01.";
RL Nucleic Acids Res. 32:977-988(2004).
DR EMBL; AE017269; AAS40708.1; --
DR TIGR; BCE1780; --
SQ SEQUENCE 465 AA; 49855 MW; E4C307E7717381D5 CRC64;

Query Match 7.5%; Score 113; DB 2; Length 465;
Best Local Similarity 22.0%; Pred. No. 23;
Matches 72; Conservative 53; Mismatches 116; Indels 86; Gaps 15;

QY 6 AEOTVEVKSIAETADGALDLYNKYLDQV---IPWKTFTDETIKELSRFK-----Q 52
DB 184 AKATI-TAKDAFDAKADSDAAKKEIDAAAKOTPSKNDAAQLAKEYIEAKATLATLKPTDA 242

QY 53 EYSQEASVL-----VGDIKVLIMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNE 106
DB 243 TYAAKAAELDAATTAALNDNAKVLVDGYEKKLTTTKTEA-----BYTA 285

```

Query Match 7.5%; Score 113; DB 1; Length 956;  
 Best Local Similarity 24.4%; Pred. No. 53;  
 Matches 68; Conservative 45; Mismatches 78; Indels 88; Gaps 15;

QY 107 KKSAQKDIILIRILDGVKKLNEAQSLLTSSQSF--NNASGKLLA---LDSQLT-----N 157  
 DB 286 AKGOSTK-----STAAADLVTKYETAKSNALGNDIAKEYLEAKTAYEANKN 331  
 QY 158 DFGEKSSYFQSDVRIRKEAYAGAAAGIVAGPGLLIISYSIA--AGVIEGKL--IPELNN 213  
 DB 332 DISSKSR--FEAAETELNKDITANKAAKV-----LVETYEKATGTTGTEKSLVAVDKIDE 384  
 QY 214 RLKTVQNFSSLSNTVKQAKNDIDAAKLLATBIAAIGEI-----KTETETTRFRVVDYD 267  
 DB 385 ALKTIADNRATLGATLNRLLDFNVNVLKASQASMASASQIEDADMAKENSEWTKP----- 439  
 QY 268 DLMSLKGAAGKAKMINTCNEYQQRHGK 294  
 DB 440 ----KILNEAGISMSQANQTPQMVSK 462

RESULT 46  
 YEF3 YEAST  
 ID YEF3 YEAST STANDARD; PRT; 956 AA.  
 AC P32618;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DE Hypothetical 106.1 kDa protein in GLY1-GDAI intergenic region.  
 GN OrderedLocNames=YEL043W; ORFNames=SYGP-ORF14;  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97131264; PubMed=9169868;  
 RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,  
 RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,  
 RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,  
 RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,  
 RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,  
 RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S.,  
 RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;  
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome V";  
 RT Nature 387:78-81(1997).  
 RN [2]  
 RN FIBRONECTIN TYPE III DOMAIN  
 RP MEDLINE=97148176; PubMed=8994808;  
 RX Bateman A., Chothia C.;  
 RA "Fibronectin type III domains in yeast detected by a hidden Markov  
 model";  
 RT Curr. Biol. 6:1544-1546(1996).  
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U18779; AAB64999.1; --  
 DR PIR; S30834; S30834.  
 DR GERMOnline; I39047; --  
 DR SGD; S0000769; YEL043W.  
 DR GO; GO:0005783; C:endoplasmic reticulum; IDA.  
 DR InterPro; IPR003361; FN.III.  
 DR InterPro; IPR008957; FN.III-like.  
 DR SMART; SM00060; FN3; 1.  
 KW Hypothetical protein.  
 FT DOMAIN 35 125  
 FT SEQUENCE 956 AA; 106132 MW; 3F78B09A0FCA03AF CRC64;

QY 38 KTDFTIKELSRFKQYSQFASVLDGDIKVLMD-----SQDKYFEATQTVVWCGVVQ 92  
 DB 272 KSLKSNIKSL-----ENSKLLTDLKEKLNKIDKSKESKISQWRNDQMKSQEDTE 322  
 QY 93 LLS-----AVILLFQYNEKASAKQDI-----LIRILDDGVKKLNEAQSLLTSS--S 138  
 DB 323 LLSKDTIKKFKYKLLNESNASVANINKEISLQNEISKMEESNKLNLASKSLTISIVN 382  
 QY 139 QSFNN-----ASGKLLALDSQITVDFE--EKSSYF-----QSOVDRIKAEYA 179  
 DB 383 ANVENDKPIASGELSAVLKEL--NDFLEKNGFUSNAGEEFLSKLNADSSLLIKMKOE--- 438  
 QY 180 GAAAGIVAGPGLLIISYSIAA-----GVIEGKLIP--ELNNR--LKT-----V 218  
 DB 439 -----LSIDQLEANWKLQRSNLLKKISALENQFNWNLNRLNLTKLAVQPY 486

219 QNFTLSAT-----VKQAKNDIDAAKLLATBIAAIGE 253  
 487 KNGDSLAATNSNNSAENKSSGSIQLPLNNMSTGSI 525

RESULT 47  
 RA50 METJA STANDARD; PRT; 1005 AA.  
 ID RA50 METJA  
 AC Q58718;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE DNA double-strand break repair rad50 ATPase.  
 GN Name=rad50; OrderedLocNames=WJ1322;  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Adams M.D., Reich C.I.,  
 RA Karlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Glodek A.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Nguyen D.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Sadow P.W., Hanna M.C.,  
 RA Usterback T.R., Kelley J.M., Peterson J.D., Kaine B.P., Borodovsky M.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii";  
 RT Science 273:1058-1073(1996).  
 CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The  
 CC rad50/mre11 complex possesses single-strand endonuclease activity  
 CC and ATP-dependent double-strand-specific 3'-5' exonuclease  
 CC activity. Rad50 provides an ATP-dependent control of mre11 by  
 CC unwinding and/or repositioning DNA ends into the mre11 active site  
 CC (By similarity).  
 CC -!- COPACTOR: Binds 1 zinc ion per heterotetramer (By similarity).  
 CC -!- SUBUNIT: Heterotetramer composed of two mre11 subunits and two  
 CC rad50 subunits (By similarity).  
 CC -!- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U67572; AAB99331.1; --  
 DR

```

DR PIR: A64465; A64465.
DR HSSP; P58301; 1F2T.
DR TIGR; MJ1322; -.
DR HAMAP; MF_00449; -.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR007517; Rad50_zn_hook.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF04423; Rad50_zn_hook; 1.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR01725; phage_HK97_gp10; 1.
KW ATP-binding; Coiled coil; Complete proteome; DNA repair; Hydrolase;
KW Zinc.
FT NP_BIND 32 39 ATP (By similarity).
FT DOMAIN 158 849 Coiled coil (Potential).
FT METAL 502 502 Zinc (By similarity).
FT METAL 505 505 Zinc (By similarity).
SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;

Query Match 7.5%; Score 113; DB 1; Length 1005;
Best Local Similarity 19.7%; Pred. No. 57;
Matches 60; Conservative 55; Mismatches 91; Indels 98; Gaps 12;

QY 7 EQTEVVVKSATETADGALDLYNKYLDQVTPKTFDE-----TIKELSRF-----KQEYSQ 56
DB 605 DEILEDIKSOLNFKP---NFTNQYLSAVSYLVNSVDEEGIRNKEIENIVSGWNKKCRE 661

QY 57 EASVL-----VGDIKVLLMDSQDKYFEATQTVYEWGVVTVLLSAVILLPDEYNEKKAS 110
DB 662 ELNKLREDEREINRKLKLNELKNKEKE-----LIEIENRRSLKFPDKYKYLGL 710

QY 111 AQKDILRIIDGDKVLLNEAKQKSLTSSQSFNNASGKLLALD----- 152
DB 711 TEKLEEKUNIKDGLLEI-----YNICNSKILAIIDNIRKYNKEDIYIYNKK 757

QY 153 -----SQTNDPSEKSSYFQSQVDRIKREAVAGAAAGIVAGPFGILLIYSYIAAGVIEGKLI 208
DB 758 ILEVNKEINDEERISYINQKLD-----INNEE-----EHKKI 792

QY 209 PEL--NRR--LKTQVNFSTLSATVKQANKDIDAKUKLATETAAIGETITETTFYV 264
DB 793 KELYENRQELDNVREQKTEIETGIEYLLKQDVESLKRL-----KEMSNLEKEKEKLTQFV 848

QY 265 DYDD 268
DB 849 EYLD 852

RESULT 48
NAGH_CLOPE STANDARD; PRT; 1628 AA.
AC P26831;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Hyaluronoglucosaminidase precursor (EC 3.2.1.35) (Hyaluronidase) (Mu toxin).
GN Name=nagh; OrderedLocusNames=CPE0191;
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=94232189; PubMed=8177218;
RA Canard B., Garnier T., Saint-Joanis B., Cole S.T.;
RT "Molecular genetic analysis of the nagH gene encoding a hyaluronidase of Clostridium perfringens.";
```

```

RL Mol. Gen. Genet. 243:215-224 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842; DOI=10.1073/pnas.022493799;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayaishi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC -1- FUNCTION: Putative virulence factor which is likely to act on connective tissue during gas gangrene.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1->4-linkages between N-acetyl-beta-D-glucosamine and D-glucuronate residues in hyaluronate.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Contains 1 F5/8 type C domain.
CC -1- CAUTION: The partially purified protein from strain CPN50 is approximately 70 kDa smaller than the sequence indicated here.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81878; AAA23259.1; -.
DR EMBL; AP003185; BAB79897.1; -.
DR PIR; S43904; S43904.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR000421; FAS8_C.
DR Pfam; PF00404; Dockerin_1; 2.
DR Pfam; PF00754; F5_F8 type C; 3.
DR SMART; SM00231; FAS8C_1.
DR PROSITE; PS00022; FAS8C_3; 1.
KW Complete proteome; Glycosidase; Hydrolase; Signal; Toxin; Virulence.
FT SIGNAL 1 30 potential.
FT CHAIN 31 1628 Hyaluronoglucosaminidase.
FT DOMAIN 781 953 F5/8 type C.
FT VARIANT 147 147 G -> A (in strain CPN50).
FT VARIANT 172 175 KIQS -> EIKN (in strain CPN50).
FT VARIANT 250 250 V -> M (in strain CPN50).
FT VARIANT 548 548 A -> E (in strain CPN50).
FT VARIANT 558 558 D -> E (in strain CPN50).
FT VARIANT 614 614 G -> S (in strain CPN50).
FT VARIANT 944 944 I -> V (in strain CPN50).
FT VARIANT 950 950 N -> S (in strain CPN50).
FT VARIANT 979 979 T -> I (in strain CPN50).
FT VARIANT 982 982 I -> L (in strain CPN50).
FT VARIANT 1042 1042 I -> F (in strain CPN50).
FT VARIANT 1043 1628 Missing (in strain CPN50).
SQ SEQUENCE 1628 AA; 182474 MW; D4252A2512BBED69 CRC64;

Query Match 7.5%; Score 113; DB 1; Length 1628;
Best Local Similarity 23.2%; Pred. No. 1e+02;
Matches 72; Conservative 42; Mismatches 112; Indels 84; Gaps 14;

QY 27 YNKYLQDVIPWKTDFETIKELSRFKQEYSQEAS-VLVGDIKVLMLDSQDKYFEATQTVYE 85
DB 1328 YNNIYEIKIEWTVAPNIYEIITLNOEFEPFVNDLAKYDELINLSGDEY----- 1378

QY 86 WCGVTVQLLSAYILLPDEYNEKKA-----SAQKDILIRILDGCVKKNBAQKSL---LT 136
DB 1379 -----TLSSFTLKEALNEAKSLDSDNSQKKI-----DKALEKLNKAERLDLRAT 1426

QY 137 SSQSFNNAAGKLLALDSQLTND-----FSE-----KESYFQSQVDRIKRE 176
DB 1427 DFEDFN---KVLTLGNSLVEEYVTAESWALFSEVLEAANAKNKADYTDQDQINQIVID 1482

QY 177 AYAGAAGIVAGP-----FGLIISY--STAAGVIEGKLIPE-----LNNELKTVQN 220
```



Db 1483 LDASIKALVETPEVDKTNLGINOQKSLLDSEVGFVGEYHKGAKDGLTVEINKAE 1542

Qy 221 FFTSLSATVQKANDIDAOKLATEIAAIGETETTRF-----YVDYDMLSLK 276

Db 1543 VFNKEDAT---BEEINLAKESLEGALARNLSLIEBSTGDFNGNGKIDIGDLAM----- 1593

Qy 277 AAKMINTCN 286

Db 1594 VSKNIGSTN 1603

RESULT 49

ID Q71X69 PRELIMINARY; PRT; 927 AA.

AC Q71X69;

DT 05-JUL-2004 (TREMBlrel. 27, Created)

DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)

DE Membrane protein, putative.

GN OrderedLocNames=LMOF2365\_2330;

OS Listeria monocytogenes (serotype 4b / strain F2365).

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OX NCBI\_TaxID=265669;

RN [1]

RP SEQUENCE FROM N.A.

RX PubMed=15115801; DOI=10.1093/nar/gkh562;

RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,

RA Kolonay J.F., Raeko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,

RA Peterson J.D., White O., Nelson W.C., Nierman W.C., Beanan M.J.,

RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,

RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,

RA Forberger H.A., Tran B., Kathariou S., Wonderling L.D., Uhlich G.A.,

RA Bayles D.O., Luchansky J.B., Fraser C.M.,

RA "whole genome comparisons of serotype 4b and 1/2a strains of the food-

RT borne pathogen *Listeria monocytogenes* reveal new insights into the

RT core genome components of this species."

RL Nucleic Acids Res. 32:2386-2395(2004).

DR EMBL; AE017330; AAT05096.1; -

DR InterPro; IPR004089; Chmtaxis transd.

DR Pfam; PF00015; MCPsignal; 1.

DR ProDom; PD001127; Outsurface; 1.

KW Complete proteome.

SQ SEQUENCE 927 AA; 97232 MW; 9AE73C67B14685CA CRC64;

Query Match 7.4%; Score 112.5; DB 2; Length 927;

Best Local Similarity 20.7%; Pred. No. 56;

Matches 78; Conservative 43; Mismatches 135; Indels 121; Gaps 14;

Qy 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLQVLPW----KTFDETIKLSRFKQEYSQ 56

Db 160 VTKSYAEAFDKIKESGEGFAAGDSGKIKDGLVKSQEGNKTIISTNLKTLADSSLTFKD 219

Qy 57 EASVL-----VGDIK-----VLLMDSQDK 75

Db 220 GANTLEVGLKTYTDGNTAAAGDKLNDGVSTLAAGVGLKGVAAALDGGATKLSGVST 279

Qy 76 YFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASAKDILIRILDGQVKLNEAKSLL 135

Db 280 YTSQVDTL---AGINQAYTGSTALSDGLNKNGS-----VPTLASGITQLNNGOKSLA 330

Qy 136 TSSQSFNNAGSKLLA---LDSQLTN-----DFSEKSSYFQSOVDRIKEAYAGAAAGI 185

Db 331 TGLDSLVDGSKLSAGLKELDGNLTDKQKIAQLKQGMNDLQGGIDQLNKSNGVEDAA-- 388

Qy 186 VAGPFGLLIISYSTAAGVIEGKLIPELNNRL-----KTVQNFFTSLSATVKQ 231

Db 389 -----LAKQLAA---LQKSLDGLQNGLTFFIKSNANFDAEAIKSKINATAGVSAEDKQ 437

Qy 232 -----ANKDIDAOKLATEIAAIGETK-----ETETTRFYVDYDMLSLK 276

Db 438 KIIDAIOADLDKETQKSATQVATVEQLQSGLSGLDLAAIQTQVT-----ELQTG 486

Qy 277 AAKMINTCNEYQQRHG 293

Db 487 VAKISAG-----YQAVHG 499

RESULT 50

ID AAT05096 PRELIMINARY; PRT; 927 AA.

AC AAT05096;

DT 10-MAY-2004 (TREMBlrel. 27, Created)

DT 10-MAY-2004 (TREMBlrel. 27, Last sequence update)

DT 10-MAY-2004 (TREMBlrel. 27, Last annotation update)

DE Membrane protein, putative.

GN LMOF2365\_2330.

OS Listeria monocytogenes str. 4b F2365.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria;

OX NCBI\_TaxID=265669;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=4b F2365;

RX PubMed=15115801;

RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,

RA Kolonay J.F., Raeko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,

RA Peterson J.F., White O., Nelson W.C., Nierman W., Beanan M.J.,

RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,

RA Haft D.H., Selengut J., Van Aken S., Khouri H., Fedorova N.,

RA Forberger H., Tran B., Kathariou S., Wonderling L.D., Uhlich G.A.,

RA Bayles D.O., Luchansky J.B., Fraser C.M.,

RA "whole genome comparisons of serotype 4b and 1/2a strains of the food-

RT borne pathogen *Listeria monocytogenes* reveal new insights into the

RT core genome components of this species."

RL Nucleic Acids Res. 32:2386-2395(2004).

DR EMBL; AE017330; AAT05096.1; -

SQ SEQUENCE 927 AA; 97232 MW; 9AE73C67B14685CA CRC64;

Query Match 7.4%; Score 112.5; DB 2; Length 927;

Best Local Similarity 20.7%; Pred. No. 56;

Matches 78; Conservative 43; Mismatches 135; Indels 121; Gaps 14;

Qy 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLQVLPW----KTFDETIKLSRFKQEYSQ 56

Db 160 VTKSYAEAFDKIKESGEGFAAGDSGKIKDGLVKSQEGNKTIISTNLKTLADSSLTFKD 219

Qy 57 EASVL-----VGDIK-----VLLMDSQDK 75

Db 220 GANTLEVGLKTYTDGNTAAAGDKLNDGVSTLAAGVGLKGVAAALDGGATKLSGVST 279

Qy 76 YFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASAKDILIRILDGQVKLNEAKSLL 135

Db 280 YTSQVDTL---AGINQAYTGSTALSDGLNKNGS-----VPTLASGITQLNNGOKSLA 330

Qy 136 TSSQSFNNAGSKLLA---LDSQLTN-----DFSEKSSYFQSOVDRIKEAYAGAAAGI 185

Db 331 TGLDSLVDGSKLSAGLKELDGNLTDKQKIAQLKQGMNDLQGGIDQLNKSNGVEDAA-- 388

Qy 186 VAGPFGLLIISYSTAAGVIEGKLIPELNNRL-----KTVQNFFTSLSATVKQ 231

Db 389 -----LAKQLAA---LQKSLDGLQNGLTFFIKSNANFDAEAIKSKINATAGVSAEDKQ 437

Qy 232 -----ANKDIDAOKLATEIAAIGETK-----ETETTRFYVDYDMLSLK 276

Db 438 KIIDAIOADLDKETQKSATQVATVEQLQSGLSGLDLAAIQTQVT-----ELQTG 486

Qy 277 AAKMINTCNEYQQRHG 293

Db 487 VAKISAG-----YQAVHG 499

Search completed: January 5, 2005, 10:58:24

Job time : 61.6826 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:58:34 ; Search time 62.2944 Seconds  
(without alignments)  
1749.716 Million cell updates/sec

Title: US-09-993-292B-28

Perfect score: 1508

Sequence: 1 MTEIVADKTVVVKNAIETA.....TCNEYQKRGKKTLPFVPEV 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1599051 seqs, 359727711 residues

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCN\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/PCUS\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 1402  | 93.0        | 305    | 9  | US-09-993-292A-2     |
| 2          | 126   | 8.4         | 1095   | 14 | US-10-128-714-3039   |
| 3          | 126   | 8.4         | 1277   | 14 | US-10-128-714-8039   |
| 4          | 120.5 | 8.0         | 891    | 14 | US-10-369-493-10918  |
| 5          | 120.5 | 8.0         | 962    | 13 | US-10-087-192-222    |
| 6          | 119   | 7.9         | 889    | 16 | US-10-437-963-197045 |
| 7          | 118   | 7.8         | 961    | 14 | US-10-080-608A-66    |
| 8          | 118   | 7.8         | 961    | 14 | US-10-370-685-155    |
| 9          | 117.5 | 7.8         | 496    | 16 | US-10-767-701-43815  |
| 10         | 114.5 | 7.6         | 1189   | 15 | US-10-282-122A-70920 |
| 11         | 113   | 7.5         | 458    | 16 | US-10-467-421-37     |
| 12         | 112.5 | 7.5         | 1020   | 15 | US-10-282-122A-58016 |
| 13         | 112.5 | 7.5         | 1679   | 14 | US-10-369-493-22080  |

|    |       |     |      |    |                      |                    |
|----|-------|-----|------|----|----------------------|--------------------|
| 14 | 112   | 7.4 | 1875 | 14 | US-10-369-493-22285  | Sequence 22285, A  |
| 15 | 111.5 | 7.4 | 2823 | 14 | US-10-369-493-5220   | Sequence 5220, Ap  |
| 16 | 111.5 | 7.4 | 2823 | 14 | US-10-369-493-5221   | Sequence 5221, Ap  |
| 17 | 111   | 7.4 | 435  | 10 | US-09-978-309A-80    | Sequence 80, Appl  |
| 18 | 111   | 7.4 | 476  | 10 | US-09-978-309A-77    | Sequence 77, Appl  |
| 19 | 111   | 7.4 | 794  | 15 | US-10-257-377-2      | Sequence 2, Appli  |
| 20 | 111   | 7.4 | 938  | 15 | US-10-282-122A-71606 | Sequence 71606, A  |
| 21 | 111   | 7.4 | 1023 | 10 | US-09-884-696-5      | Sequence 5, Appli  |
| 22 | 111   | 7.4 | 1171 | 14 | US-10-032-585-7519   | Sequence 7519, Ap  |
| 23 | 110   | 7.3 | 1163 | 15 | US-10-282-122A-51864 | Sequence 51864, A  |
| 24 | 109.5 | 7.3 | 420  | 15 | US-10-425-114-60225  | Sequence 60225, A  |
| 25 | 109.5 | 7.3 | 660  | 13 | US-09-841-260-139    | Sequence 139, App  |
| 26 | 109.5 | 7.3 | 660  | 13 | US-10-007-693-139    | Sequence 139, App  |
| 27 | 109.5 | 7.3 | 660  | 16 | US-10-762-058-139    | Sequence 139, App  |
| 28 | 109   | 7.2 | 1005 | 14 | US-10-369-493-1061   | Sequence 1061, Ap  |
| 29 | 108.5 | 7.2 | 476  | 10 | US-09-978-309A-79    | Sequence 79, Appl  |
| 30 | 108.5 | 7.2 | 552  | 17 | US-10-739-930-6691   | Sequence 6691, Ap  |
| 31 | 108.5 | 7.2 | 647  | 15 | US-10-425-114-51665  | Sequence 51665, A  |
| 32 | 108.5 | 7.2 | 886  | 14 | US-10-369-493-1016   | Sequence 1016, Ap  |
| 33 | 108   | 7.2 | 737  | 16 | US-10-437-963-107270 | Sequence 107270, A |
| 34 | 107.5 | 7.1 | 630  | 15 | US-10-205-647A-6     | Sequence 6, Appli  |
| 35 | 107.5 | 7.1 | 631  | 10 | US-09-978-309A-48    | Sequence 48, Appl  |
| 36 | 106.5 | 7.1 | 680  | 15 | US-10-425-114-68086  | Sequence 68086, A  |
| 37 | 106.5 | 7.1 | 1999 | 14 | US-10-028-249A-107   | Sequence 107, App  |
| 38 | 106.5 | 7.1 | 1999 | 15 | US-10-107-782-107    | Sequence 107, App  |
| 39 | 106   | 7.0 | 1583 | 16 | US-10-408-765A-1635  | Sequence 1635, Ap  |
| 40 | 106   | 7.0 | 1961 | 14 | US-10-028-248A-105   | Sequence 105, App  |
| 41 | 106   | 7.0 | 1961 | 15 | US-10-107-782-105    | Sequence 105, App  |
| 42 | 106   | 7.0 | 2076 | 9  | US-09-815-242-5815   | Sequence 5815, Ap  |
| 43 | 106   | 7.0 | 2086 | 9  | US-09-815-242-5639   | Sequence 5639, Ap  |
| 44 | 106   | 7.0 | 2186 | 9  | US-09-815-242-12913  | Sequence 12913, A  |
| 45 | 106   | 7.0 | 5795 | 9  | US-09-815-242-12610  | Sequence 12610, A  |
| 46 | 105.5 | 7.0 | 459  | 15 | US-10-424-599-248508 | Sequence 248508, A |
| 47 | 105.5 | 7.0 | 742  | 15 | US-10-282-122A-53523 | Sequence 53523, A  |
| 48 | 105.5 | 7.0 | 815  | 16 | US-10-437-963-191043 | Sequence 191043, A |
| 49 | 105.5 | 7.0 | 873  | 14 | US-10-369-493-11332  | Sequence 11332, A  |
| 50 | 105.5 | 7.0 | 1312 | 14 | US-10-393-602-148    | Sequence 148, App  |

ALIGNMENTS

RESULT 1  
US-09-993-292A-2  
; Sequence 2, Application US/09993292A  
; Patent No. US20020146430A1  
; GENERAL INFORMATION:  
; APPLICANT: James B. Galen  
; APPLICANT: University of Maryland  
; TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF PROTEINS  
; FILE REFERENCE: UOPMD.007A  
; CURRENT APPLICATION NUMBER: US/09/993,292A  
; CURRENT FILING DATE: 2001-11-23  
; PRIOR APPLICATION NUMBER: 60/252,516  
; PRIOR FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Salmonella Typhi  
US-09-993-292A-2

Query Match 93.0%; Score 1402; DB 9; Length 305;  
Best Local Similarity 91.1%; Pred. No. 4.5e-107;  
Matches 276; Conservative 16; Mismatches 11; Indels 0; Gaps 0;  
Qy 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVIPHQTFDETIKLSRFKQYSQASV 60  
Db 1 MTSIFASQTVVVKNAIETADGALDLYNKYLDQVIPHQTFDETIKLSRFKQYSQASV 60  
Qy 61 LVGDIKTLMDSQKYFEATQTVVWCVATQTLAAVILLFDEYNEKASAKQDILIKVL 120

Db 61 LVGDIKULLMDSQDKYPEATQTVVWCGVVTQLLSAYILLFDYNEKSGASQAQDILIRIL 120  
Qy 121 DDGTTKLNEAQSLLVSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSQVDKIRKEAYAG 180  
Db 121 DDGVTKLNEAQSLLVSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSQVDKIRKEAYAG 180  
Qy 181 AAGVWAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLNTVVKQAKNDIDAAK 240  
Db 181 AAGVWAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLNTVVKQAKNDIDAAK 240  
Qy 241 LKLTETIAAIGEIKTETETTRFYVDYDMLSLILKEAAKWMINTCNEYQKRHGKTLFEV 300  
Db 241 LKLTETIAAIGEIKTETETTRFYVDYDMLSLILKEAAKWMINTCNEYQKRHGKTLFEV 300  
Qy 301 PEV 303  
Db 301 PDV 303

RESULT 2  
US-10-128-714-3039  
; Sequence 3039, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wenqi  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; PRIOR FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 3039  
; LENGTH: 1095  
; TYPE: PRT  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-3039

Query Match 8.4%; Score 126; DB 14; Length 1095;  
Best Local Similarity 21.4%; Pred. No. 0.26; Indels 146; Gaps 15;  
Matches 73; Conservative 49; Mismatches 73

Qy 4 IVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQYQSAASV--L 61  
Db 281 IKKEKDIEEATNALVPVDEKVDITRKVER-----PASRIAEIG--KERDSQAANVKQL 332  
Qy 62 VGDITKLMDSDQDKYFEATQTVVWCGVATQLLAAVILLFDEYNEKASQAQDILIKVLD 121  
Db 333 EKDLK--VVEKAQAQWEA-----EW----- 350  
Qy 122 DGITKLNEAQSLLVSSQSFNNASGKLLALDSQ-----LTNDFSEKSSYFQSQVDKIRKEA 177  
Db 351 -----HKTMSNKGQQLSESQDQYKMKLKEVSKRSSAEQINLDNLKRQR 394  
Qy 178 YAGAAAGVVGPFGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNFFTT----- 224  
Db 395 KTEAEA-----YN-----SLKSKFDSTEWQKLSVENDTQTLTERKS 430

Qy 225 -LSNTVKQAKNDIDAAK--LTTETIAAIGEIKTETETTRFYVDYDMLSLILKEA--- 277  
Db 431 ALNDTVKTTSEIKDRKKELNALTSRLRISQMRTELE-----EKQVVVLKKLEADDG 484  
Qy 278 -----AKWMINT-----CNEYQKRHGK 294  
Db 485 KKQTERELRAKELISTLKRIFPFGVGRVSDLCRPKQKKYAE 525

RESULT 3  
US-10-128-714-8039  
; Sequence 8039, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wenqi  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; PRIOR FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 8039  
; LENGTH: 1277  
; TYPE: PRT  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-8039

Query Match 8.4%; Score 126; DB 14; Length 1277;  
Best Local Similarity 21.4%; Pred. No. 0.32; Indels 146; Gaps 15;  
Matches 73; Conservative 49; Mismatches 73

Qy 4 IVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQYQSAASV--L 61  
Db 323 IKKEKDIEEATNALVPVDEKVDITRKVER-----PASRIAEIG--KERDSQAANVKQL 374  
Qy 62 VGDITKLMDSDQDKYFEATQTVVWCGVATQLLAAVILLFDEYNEKASQAQDILIKVLD 121  
Db 375 EKDLK--VVEKAQAQWEA-----EW----- 392  
Qy 122 DGITKLNEAQSLLVSSQSFNNASGKLLALDSQ-----LTNDFSEKSSYFQSQVDKIRKEA 177  
Db 393 -----HKTMSNKGQQLSESQDQYKMKLKEVSKRSSAEQINLDNLKRQR 436  
Qy 178 YAGAAAGVVGPFGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNFFTT----- 224  
Db 437 KTEAEA-----YN-----SLKSKFDSTEWQKLSVENDTQTLTERKS 472  
Qy 225 -LSNTVKQAKNDIDAAK--LTTETIAAIGEIKTETETTRFYVDYDMLSLILKEA--- 277  
Db 473 ALNDTVKTTSEIKDRKKELNALTSRLRISQMRTELE-----EKQVVVLKKLEADDG 526  
Qy 278 -----AKWMINT-----CNEYQKRHGK 294  
Db 527 KKQTERELRAKELISTLKRIFPFGVGRVSDLCRPKQKKYAE 567

## RESULT 4

US-10-369-493-10918  
; Sequence 10918, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 10918  
; LENGTH: 891  
; TYPE: PRT  
; ORGANISM: Ferroplasma acidimanus  
US-10-369-493-10918

Query Match 8.0%; Score 120.5; DB 14; Length 891;  
Best Local Similarity 21.9%; Pred. No. 0.56; Mismatches 47; Indels 55; Gaps 10;  
Matches 64; Conservative 47;  
QY 8 KTVVVK-NAETADGALDLYNKYLDQVPMQTFDETIKELSRFKQYSQAASVLVG 63  
DB 522 KSLTEKKELENTIENASKGHELVLSAEN-ENLEKAIKLOQYENYIRVSSIIG 578  
QY 64 DIKTLMD-SQDKYFATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQKOLIKVLD 121  
DB 579 -MDPGAIIKEIEAAGESFTTFKRNINELLSQLGFPVEYNEQYNTGKISSEINRLK 632  
QY 122 DGIKLEAQSLLVSSQSFNNASGKLALDLSQTLNDFSEKSSYFQSQVDKIRKEAAGA 181  
DB 633 TEVERSEMSKLESIKDEIENRKSIEGLRIENKOSAMHOY-DGIDSQ----- 682  
QY 182 AAGVAGPFGLLIISYSIAAGVVEGKLIPELKNKLS-VONFFTLNNTVKQA-NKIDID 237  
DB 683 -AGNIESRYKSAQENIKMTLVDSYTERIEETEENAKNLEQD 724  
QY 238 AAKULTE-IAAIGEIKTETETTRFYVDYDDMLSLKLLKAAKQKQINTCNEY 288  
DB 725 AEKYKTREAITSLTKI-----REAFDYNIGIQSIIRKQDASASMTNLTRY 769

## RESULT 5

US-10-087-192-222  
; Sequence 222, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 222  
; LENGTH: 962  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-087-192-222

Query Match 8.0%; Score 120.5; DB 13; Length 962;  
Best Local Similarity 23.7%; Pred. No. 0.62; Mismatches 136; Indels 51; Gaps 13;  
Matches 74; Conservative 51;  
QY 7 DKTVEVVKNAIETADGALDLY-NKYLDQVPMQTFDETIKELSRFKQYSQAASVLVGD 65  
DB 635 DKKEEVKKTLEQHDNIVTHYKNMIREQDLQLELRQOVSTLKQNSQLQTAFTVQVSO 694  
QY 66 K-----TLIM-----DSQDKYFATQTVYEWCGVATQLLAAAYILLFDEYNEKKA 110  
DB 695 QOHKQDYNLLKIQLGKDNQHQSGYSEGAQ-----MNGIQPEIGR---LREIEBELKRNQE 747  
QY 111 -AQKDLIKVLDDGTTKLENAOKSLVLS--SQSFNNASGKLALDLSQTLNDFSE 161  
DB 748 LLOSQLTEKDSMIENMKSSQTSQNEQSSAIVSARDSEQVAELKQELATLKSQI-NQSOV 806  
QY 162 KSSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKL-----IPELKNKL 215  
DB 807 EITKQTEKQELQKTEAFKSVQVQETETIATKTTD-VEGRLSALLQETKELKNEI 864  
QY 216 KSVQNFPTLSNTVKQANKOIDAOKLKLTTETIAAIGIKTETETTRFYVDYDDMLSLK 275  
DB 865 KALSEERTAIKEQLDSSNSTIAI-----LQTE-----KDKLELEITDSKQEQDILLV-LLA 914  
QY 276 EAAKQMINTCNE 287  
DB 915 DQOKILSLKNK 926  
RESULT 6  
US-10-437-963-197045  
; Sequence 197045, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 197045  
; LENGTH: 889  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_92840C.1.pep  
US-10-437-963-197045

Query Match 7.9%; Score 119; DB 16; Length 889;  
Best Local Similarity 21.1%; Pred. No. 0.74;  
Matches 74; Conservative 74; Mismatches 123; Indels 80; Gaps 18;  
QY 1 MTEIVADKTVVVKNAIETADGALD-----LYNKYLDQVPMQTFDETIK-----EL 47  
DB 262 IAEVNAEK-VEILSSEVVRUKGLDSDTAESSEKSNRETELV--KNLESEVSVLKGKLEE 318  
QY 48 SRFKQYSQAASVLVDIKTLMDSDQKYFATQTVVW---CGVATQLLAAAYILLFDEY 104  
DB 319 ARIIEERLAETEKLIEELKSEVADAKKAESEARQLPEWKHKAGLLEMLEA-VLSDKF 377  
QY 105 -NEKKAQAOKDILIKVLDGTTKLENAOKSLVSSQSFNNASGKLALD---SQTNDPFS 160  
DB 378 KGESLASTTEE-----LGKIQSALQDRESIEVLKGTTLTALEVARLLADVN 425  
QY 161 EKSYF-----SQVDKIRKEAYAGAAAGVAGPFGLLIISYSIAAGV-VEG----- 205

Wed Jan 5 14:16:57 2005

```

; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 155
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-370-685-155

Query Match      7.8%; Score 118; DB 14; Length 961;
Best Local Similarity 22.9%; Pred. No. 1;
Matches 72; Conservative 54; Mismatches 138; Indels 50; Gaps 12;

Db 426 ESNEQFASQVEFGLQTTIDVLRKLEAAEAASEA-----LNNEKAANKVIEGLTEE 479
Qy 206 --KLIPEL-----KNKLKSVQNFHTLNTVKQANKDIDAAKLLTTE-----IAAI 250
Db 480 NVKLISELNETRDEEKEKRAVEDLTAALS---EESDKAKEAHERYLSKEDDHEHALAQI 536
Qy 251 GEITETETTR---FVYVDLMLSLLEAKAKMINTCNEYQKR-HGKKT 296
Db 537 GDLKALKSKSYEVVLMDEANYDITCLRNVDKLEAEVNVKRECESEKET 587

RESULT 7
US-10-080-608A-66
; Sequence 66, Application US/10080608A
; Publication No. US2003019956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-080-608A-66

Query Match      7.8%; Score 118; DB 14; Length 961;
Best Local Similarity 22.9%; Pred. No. 1;
Matches 72; Conservative 54; Mismatches 138; Indels 50; Gaps 12;

Qy 7 DKTVEVVKNAIETADGALDLY-NKYLDPVWPQTFDTIKELSRFKQYEQYSAASVLVGD 65
Db 635 DKKEEVKKTLEQHDHSIVTHYKNMIREQDLQLEELKQOIISTLKQNEQLQTAVTQVQS 694
Qy 66 K-----TLMDSDQK--YFEATQTVYEWGCVATQLLAAAYILLFDEYNEKKAS-- 110
Db 695 QQHKKDQVNLKVLQKDSHQGPYTDGAQ-----MNGVQPEEISR--LREEIEELKSN 747
Qy 111 -----AQKDILIKVLDDGITK--LNEAQKSLVSSQFNNSAGKLLALDSQLTNDFSE 162
Db 748 LLOSQAELKDSLIENLKSSQLSPGTNEQSATAGDSEQIAELKQELATLKSQ--NSQSV 806
Qy 163 SSVFQSQVDKIRKEAYAGAAAGVAGPFGIISYSIAAGVVEGKL-----IPELKNK 216
Db 807 ITKLQTEKQELLQKTEAFKSAFPVGESETVIATKTTD--VEGRLSALLQETKELKNE 864
Qy 217 SVQNFFTLNTVKQANKDIDAAKLLTTEIAAIGIKTETETTRFYVDYDMLSLKKE 276
Db 865 ALSEERTAIKEQLDSSNS-----TIALQNEKNKLEVDITDSKKEQDOLLV-LLAD 914
Qy 277 AAKMINTCNEYQK 290
Db 915 QDQKIFSLKNKLE 928

US-10-080-608A-66

Query Match      7.8%; Score 118; DB 14; Length 961;
Best Local Similarity 22.9%; Pred. No. 1;
Matches 72; Conservative 54; Mismatches 138; Indels 50; Gaps 12;

Qy 7 DKTVEVVKNAIETADGALDLY-NKYLDPVWPQTFDTIKELSRFKQYEQYSAASVLVGD 65
Db 635 DKKEEVKKTLEQHDHSIVTHYKNMIREQDLQLEELKQOIISTLKQNEQLQTAVTQVQS 694
Qy 66 K-----TLMDSDQK--YFEATQTVYEWGCVATQLLAAAYILLFDEYNEKKAS-- 110
Db 695 QQHKKDQVNLKVLQKDSHQGPYTDGAQ-----MNGVQPEEISR--LREEIEELKSN 747
Qy 111 -----AQKDILIKVLDDGITK--LNEAQKSLVSSQFNNSAGKLLALDSQLTNDFSE 162
Db 748 LLOSQAELKDSLIENLKSSQLSPGTNEQSATAGDSEQIAELKQELATLKSQ--NSQSV 806
Qy 163 SSVFQSQVDKIRKEAYAGAAAGVAGPFGIISYSIAAGVVEGKL-----IPELKNK 216
Db 807 ITKLQTEKQELLQKTEAFKSAFPVGESETVIATKTTD--VEGRLSALLQETKELKNE 864
Qy 217 SVQNFFTLNTVKQANKDIDAAKLLTTEIAAIGIKTETETTRFYVDYDMLSLKKE 276
Db 865 ALSEERTAIKEQLDSSNS-----TIALQNEKNKLEVDITDSKKEQDOLLV-LLAD 914
Qy 277 AAKMINTCNEYQK 290
Db 915 QDQKIFSLKNKLE 928

US-10-080-608A-66

Query Match      7.8%; Score 117.5; DB 16; Length 496;
Best Local Similarity 20.2%; Pred. No. 0.45;
Matches 68; Conservative 49; Mismatches 134; Indels 85; Gaps 10;

Qy 32 DQVWPQTFDE-----TKELSRFKQYEQYSAASVLVGDITKLMDSDQKYEATQTVYEW 87
Db 76 DKATDHQOIIEESLRSITSEL-KVVQEALELSKSVADLEQKLSAQDADISKLTEELNHC 134
Qy 88 GVATOLLAAAYILLFDEYNEKKASAKOILIKVLDDGITKLENAOKSLVSSQFNNSAGK 147
Db 135 SSESLEKTKLLETETLTALLEELOAKLL--SLQEMETKLEDEQSKGRQTSAALEKONGQ 192
Qy 148 LLALDSQLTN-----DFSEKSSYFQSQVDKIRK----- 175

```

```
Db 193 LIVLQABLDNLKDNFTLQSLADLNKFSKSNLHQAEDLAKAQVILSEALSQKEEL 252
Qy 176 -----EAYAGAAGVWAGPFGLLIISYSI-----AAGVVEGKLIPELK----- 212
Db 253 ELNLSLSEQHESKA-----FCENASQKILELEAQVHAMHAAEALNLELKEAASVK 306
Qy 213 ---NKLKSVQNFPTLNTVQKANDIDAAKLKUTTEIAAIGEIKTETETTRFYVD-YDD 268
Db 307 AAENKSDLEQQLSEIENKLVASSEEIELKRIQOEAAVSAERGMLQETMTSVEGYKE 366
Qy 269 LML-----SILKEAAKMTNCTNEYOKR 291
Db 367 KITELQSLSSSVSKNQLLEQVEKELDKCSEHQEQ 402

RESULT 10
US-10-282-122A-70920
; Sequence 70920, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70920
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70920

Query Match 7.6%; Score 114.5; DB 15; Length 1189;
Best Local Similarity 20.1%; Pred. No. 2.6;
Matches 67; Conservative 45; Mismatches 96; Indels 125; Gaps 13;

Qy 5 VADTKVEVVK-----NAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFKOEYSQAA 58
Db 161 IIEESAGLVKTKKKAESIQKLDHTEDNLNRVEDILYDLGRVPLKEEAIAKEYKQLS 220
```

```
Qy 59 S-----VLVGDIKTLLMDSQDKYFEATQTVYEW-----GVATQLLAA 96
Db 221 KEMEQSIVITVSDI-----DHYTEDNQLRDLRNLHLKSOQAEKEGQQAQINQLLQR 272
Qy 97 YILLFDEYNEKKASAKQDILIKVLDDGIGITKLEAQAQKLLVSSQSFNNASGKLLALDSOLT 156
Db 273 Y-----KGRQON-----DYDIEKLN-----YELVKATENYEQLSGKLNVLZEERKK 313
Qy 157 ND-----FSEKSSYFQSOVDKIRKEAYAGAAAGVWAGPFGLLIISYSIAAGVVEGKLPIE 210
Db 314 NQSETNARYEELDNLESQIDSINKV-----AQNEKLLAD 349
Qy 211 LKN-----KLSKVNQFTTLLSNTVQKANKDI----- 236
Db 350 LKNKQKQLNKEVOELESLLYVSDQHDQKLEEKNSYITLMSQESQSDVNNDIRPLEHTINE 409
Qy 237 -DAAKLKLTE-IAAIGEIK-----TETETTRFY 263
Db 410 NEAKSKRLDSRLVEAFNQLKDIOQNIITQTNKAY 442

RESULT 11
US-10-467-421-37
; Sequence 37, Application US/10467421
; Publication No. US20040116665A1
; GENERAL INFORMATION:
; APPLICANT: Berthet, Francois-Xavier Jacques
; APPLICANT: Denoel, Philippe
; APPLICANT: Neyt, Cecile Anne
; APPLICANT: Poolman, Jan
; APPLICANT: Thonnard, Joelle
; TITLE OF INVENTION: Vaccine Composition
; FILE REFERENCE: B45259
; CURRENT APPLICATION NUMBER: US/10/467,421
; CURRENT FILING DATE: 2003-08-08
; PRIOR APPLICATION NUMBER: PCT/EP02/01361
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: GB 0103171.5
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-467-421-37
```

```
Query Match 7.5%; Score 113; DB 16; Length 458;
Best Local Similarity 20.4%; Pred. No. 0.95;
Matches 68; Conservative 57; Mismatches 131; Indels 78; Gaps 12;

Qy 6 ADKTVVVKNAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFKOE-----YSQA 57
Db 102 ATEKLEVAKEATQ-----DKVEKTQSLVEDIKDKAQSLOEADAADTVEALKQA 148
Qy 58 ASVLVGDIKTLLMDSQDKYFEATQTVYEW-----GVATQLLAAYILLFDEYN--- 105
Db 149 AS---DKVETTKAEASQSLKDDATQT-FESAKQAVEGKVEAIKEQVLQVDVSLKDDTDQDN 204
Qy 106 ----EKKASAKQDILIKVLDDGIGITKLEAQAQKLLVSSQSFNNASG-----KLLALD--SQ 154
Db 205 TDQDQEKQTLKQAVQAATAAKRKVEDVDVDVHTTESFKNTASEKIDIKQAADVTEE 264
Qy 155 LTNDSEKSSYFQSOVDKIRKEAYAGAAAGVWAGPFGLLIISYSIAAGVVEGKLPIELKNK 214
Db 265 VKSLSQKADALSSGSELKQTAQT-ANDAITAQAQAVVSGVAAADSQAQSAKDK 323
Qy 215 LKSVQNFPTLNTVQKANKDIDAAKLKUTTEIAAIGEIKTETETTRFYVDYDDMLSL 274
Db 324 L-----NQLFEQKGSALDEKQVQELGEYGA-TEKINAVSEYVDLATQVI 365
Qy 275 KEAKGMINTC-----NEYQKRHGKKT 297
```

Db 366 KEEAQLQTNQAESLQAAGAAGEYDATHEDKGL 399

RESULT 12  
US-10-282-122A-58016  
; Sequence 58016, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 58016  
; LENGTH: 1020  
; TYPE: PRT  
; ORGANISM: Enterococcus faecium  
US-10-282-122A-58016

Query Match 7.5%; Score 112.5; DB 15; Length 1020;  
Best Local Similarity 20.5%; Pred. No. 3.1;  
Matches 64; Conservative 46; Mismatches 105; Indels 97; Gaps 11;  
QY 18 ETADG---ALDLNKKYLDQVPMQTFDETIKLSRQKQYSA-ASVLVGDITKLIMDSQ 73  
DB 183 ETEDNLSRVODITHELEEQTFPLAQSEAKKEFLRLKETLTQTDVSLMVAEIKTKAD-- 240  
QY 74 DKYFEATQTVYECGVATQLAAAYILLFDEYNKKASAKDILIKVLDDGITYKLNAQKS 133  
DB 241 -----NDNKQAQL-----AKFNLEFLGLSESI-----QEQES 267  
QY 134 LTVSSQFNNAASKLLALDSQTLNDRSEKSSYFQSOVDKIRKAYAGAAAGVAGPGLI 193  
DB 268 ILAKQRKENAQADRLFEKNQOVLDDSEKLUKQTEGQKQVLTQRT----- 311  
QY 194 ISYSIAAGVVEGKLIPELKNKLSVQNFPTLNTVTKQA-----NKDIDAAKLKLTTEI 247  
DB 312 -----KHTQKSSQEQYQSLAEAAQKKVKHFEKLQESLMKAAAEKETEI 353

QY 248 --AAIGETKETE-----TTFYVDYDDMLMLSLKEAAKMMINTCNEYQKRH----- 292  
DB 354 QKAEANLIKTOELEKYQKSTKELLAEURDQYVDLMQEOA--AVGNELKYLERYQIQTETA 411  
QY 293 -GKKTLEFVPEV 303  
DB 412 KSKQTLAKQSEV 423  
RESULT 13  
US-10-369-493-22080  
; Sequence 22080, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 22080  
; LENGTH: 1679  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-22080

Query Match 7.5%; Score 112.5; DB 14; Length 1679;  
Best Local Similarity 20.6%; Pred. No. 6;  
Matches 51; Conservative 44; Mismatches 67; Indels 85; Gaps 10;  
QY 39 TPDETIKELSR-----FKQYSAASVLVGDITKLIMDSQKYFEATQTVYECGVATQLL 94  
DB 1134 SLEKTIINDLQRTLTSEKY-QCSAVIIDFKDITK-----EVTQV----- 1173  
QY 95 AAVILLFDEYNKKASAKQDI-----LIKVLDDGITKLNEAKSLVSSQSFNNAS 145  
DB 1174 -----NILKENNAIILQKSLKNVTEKNREIYKQLNDROBEISRLQRLDTQTEQVSINS 1226  
QY 146 GKLIALDSQTL-----NDFSEKSSYFQSQ-VDKIRKAYAGAAAGVAGPGLIISYSI 198  
DB 1227 NKILVSEMEQCKQRYQDLSQOQKDAQKDIKLTNE----- 1264  
QY 199 AAGVVEGKLIPELKNKLSVQNFPTLNTV-----KQANKDIDAALK-----LTTEIAAIG 251  
DB 1265 -----ISDLKGKLSAENANADLENKFNRLKQAEKLDASKKQQAALTNELNELK 1315  
QY 252 EIKTETE 258  
DB 1316 AIKDKLE 1322

RESULT 14  
US-10-369-493-22285  
; Sequence 22285, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 22285  
; LENGTH: 1875  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-22285

Query Match 7.4%; Score 112; DB 14; Length 1875;  
Best Local Similarity 19.7%; Pred. No. 7.6;  
Matches 68; Conservative 66; Mismatches 122; Indels 90; Gaps 15;

Qy 2 TEIVADTKVTVVVKNAIETADGALDLYNK---YLDQVI-----PWQTFDETIK----- 45  
Db 685 SRILAEERFKLLSNTLDLTKAENDQLRKRFDYLTQTLKQDSKTHETLNEYVSCSKLSI 744  
Qy 46 ---ELSRFKQY-----SQAAVLYGDIKTLLWDSQDKYFEATQT 82  
Db 745 VETELLNLKEEQKLVHLEKVLKQELNKLSPKDSLRIMVTQTLQKREDLLEETRKS 804  
Qy 83 VYECGVATQLLAAYILLFDEYNE---KKASAQKDLIKVL-DDGITTILNEAQSLLV 136  
Db 805 CQK-----KIDELDALSELKKTQSKDHIIKQLEEDNNSNIEWQNKIEA 850  
Qy 137 SSQSFNNASGKLALDSQLTN---DFSEKSSYFQSQVDKIRKEAYAGAAAGVVGPFGL 192  
Db 851 LKQDYEV---ITSVDSKQTDIEKLVKVSLEKEIEEDKIRLHTY-----NVMD---ET 899  
Qy 193 IISYSIAAGVVEGKL-IPELKNKLKSVQNFPTLNTVVKQANKDIDAAKLKLTTTIAAIG 251  
Db 900 INDDSLRLEKSKINLTDAYSQIKYKDYETTSQSLQNTSKLDESFKDFTNQIKNLT 959  
Qy 252 EIKTETTRFYVDYDMLSLLEAAKMKMTNCNE--YORHGKK 295  
Db 960 DEKTSLED-----KISLLKE---QMFNLNLELDLQKGNEX 992

RESULT 15  
US-10-369-493-5220  
; Sequence 5220, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 5220  
; LENGTH: 2823  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-5220

Query Match 7.4%; Score 111.5; DB 14; Length 2823;  
Best Local Similarity 21.3%; Pred. No. 15;  
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

Qy 25 DLYNKYLDQVTPWOT-FDETIKLSRFKQSVQAASVLVG-----DIKTLMD 72  
Db 1881 DLKNR-IDVLEQWMDYRETIYDVSKDADAERMSLVGKRINRYKEVSEIEKLARVEA 1939  
Qy 73 QDK-----YFEATQTV-----YEWCGVATQLLAAYILLFDEY 104  
Db 1940 EDQIAYSRSNIEKARSEELNMFEDEKINWTLAELPDLVEQCONITLL---YSOLIDEY 1996

Qy 105 NE-----KKASAQKDLIKVLDDGTTKLEAOKSLVSS-----QSFNNAS--- 145  
Db 1997 DEEYVQTAGRHAELKLEVQAQ---KIVDRFVDTTETENPLKASHAYENIVEALKNATEAV 2053

Qy 146 -----GKLALDSQLTNDFSEKSSYFQSQVDKIRKEAYAGAAAGVVGPFGLIIS 195  
Db 2054 DSAEAASEAVSKMLGSGESGDAEES---LRSQLEKLNES-----S 2094

Qy 196 YSIAAGVVEGKLIPELKNKLKSVQNFPTLNTVVKQANKDIDAAKLKLTTTIAAIGIKT 255  
Db 2095 LSNVDNSNAVKIVEELKKEKKD-----LTDRLGHLN-----ELK-TSIVKRLGVKN 2140

Qy 256 ETETTRFYVDYDMLSLLEAAKMK-INTCNEYQKRGHKK 296  
Db 2141 EASS---WDDKDRMHSILKNGAKTAHERSANVKESEGIKT 2179

RESULT 17  
US-10-369-493-5220  
; Sequence 5220, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 5221  
; LENGTH: 2823  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-5221

Query Match 7.4%; Score 111.5; DB 14; Length 2823;  
Best Local Similarity 21.3%; Pred. No. 15;  
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

Qy 25 DLYNKYLDQVTPWOT-FDETIKLSRFKQSVQAASVLVG-----DIKTLMD 72  
Db 1881 DLKNR-IDVLEQWMDYRETIYDVSKDADAERMSLVGKRINRYKEVSEIEKLARVEA 1939  
Qy 73 QDK-----YFEATQTV-----YEWCGVATQLLAAYILLFDEY 104  
Db 1940 EDQIAYSRSNIEKARSEELNMFEDEKINWTLAELPDLVEQCONITLL---YSOLIDEY 1996

Qy 105 NE-----KKASAQKDLIKVLDDGTTKLEAOKSLVSS-----QSFNNAS--- 145  
Db 1997 DEEYVQTAGRHAELKLEVQAQ---KIVDRFVDTTETENPLKASHAYENIVEALKNATEAV 2053

Qy 146 -----GKLALDSQLTNDFSEKSSYFQSQVDKIRKEAYAGAAAGVVGPFGLIIS 195  
Db 2054 DSAEAASEAVSKMLGSGESGDAEES---LRSQLEKLNES-----S 2094

Qy 196 YSIAAGVVEGKLIPELKNKLKSVQNFPTLNTVVKQANKDIDAAKLKLTTTIAAIGIKT 255  
Db 2095 LSNVDNSNAVKIVEELKKEKKD-----LTDRLGHLN-----ELK-TSIVKRLGVKN 2140

Qy 256 ETETTRFYVDYDMLSLLEAAKMK-INTCNEYQKRGHKK 296  
Db 2141 EASS---WDDKDRMHSILKNGAKTAHERSANVKESEGIKT 2179

```
US-09-978-309A-80
; Sequence 80, Application US/09978309A
; Publication No. US20030100490A1
; GENERAL INFORMATION:
; APPLICANT: Cruz, Tony
; APPLICANT: Pastrak, Aleksandra
; APPLICANT: Turley, Eva A.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to
; TITLE OF INVENTION: Injury and Other Proliferating Cell Disorders Regulated by
; TITLE OF INVENTION: Hyaladerin and Hyaluronans
; FILE REFERENCE: 033352-010
; CURRENT APPLICATION NUMBER: US/09/978,309A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 09/685,010
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 09/541,522
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/127,457
; PRIOR FILING DATE: 1999-04-01
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-978-309A-80

Query Match          7.4%; Score 111; DB 10; Length 435;
Best Local Similarity 20.5%; Pred. No. 1.3; Mismatches 63; Indels 62; Gaps 15;
Matches 60; Conservative 63;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVPMQTFDETTELKELSRFKQVYSQAASV 60
DQ 99 LDNLLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKTLLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLDFEYNEKKAQAQKDLIKVL 120
DQ 156 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 205
QY 121 DDGITKLNKAQKSLI-VSSQSPFNASGKLAL-DSQLTN-DPSEKSSYFQSQVDKIRKEA 177
DQ 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLNLTLQEKVMAEKSVEDVQOOI 265
QY 178 YAGAAAGVAVGPFGLIISYIAAGVVEGKLIPELKNK-----LKSVQNF-----TTLS 226
DQ 266 LTAESTNQ-----EVA-----RMVQDLQNRSTLKEEIKETITSSFLEKITDLK 308
QY 227 NTVKQANKDI-----DAAKLKLTEIAAIGEIKETITETTFYVDYDDL 269
DQ 309 NQLRQODEDFRKQLEBKGRKTAENVMTELT-----MEINKWRLLYEEL 353

RESULT 18
US-09-978-309A-77
; Sequence 77, Application US/09978309A
; Publication No. US20030100490A1
; GENERAL INFORMATION:
; APPLICANT: Cruz, Tony
; APPLICANT: Pastrak, Aleksandra
; APPLICANT: Turley, Eva A.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to
; TITLE OF INVENTION: Injury and Other Proliferating Cell Disorders Regulated by
; TITLE OF INVENTION: Hyaladerin and Hyaluronans
; FILE REFERENCE: 033352-010
; CURRENT APPLICATION NUMBER: US/09/978,309A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 09/685,010
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 09/541,522
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/127,457
; PRIOR FILING DATE: 1999-04-01
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-978-309A-80

Query Match          7.4%; Score 111; DB 10; Length 435;
Best Local Similarity 20.5%; Pred. No. 1.3; Mismatches 63; Indels 62; Gaps 15;
Matches 60; Conservative 63;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVPMQTFDETTELKELSRFKQVYSQAASV 60
DQ 99 LDNLLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKTLLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLDFEYNEKKAQAQKDLIKVL 120
DQ 156 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 205
QY 121 DDGITKLNKAQKSLI-VSSQSPFNASGKLAL-DSQLTN-DPSEKSSYFQSQVDKIRKEA 177
DQ 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLNLTLQEKVMAEKSVEDVQOOI 265
QY 178 YAGAAAGVAVGPFGLIISYIAAGVVEGKLIPELKNK-----LKSVQNF-----TTLS 226
DQ 266 LTAESTNQ-----EVA-----RMVQDLQNRSTLKEEIKETITSSFLEKITDLK 308
QY 227 NTVKQANKDI-----DAAKLKLTEIAAIGEIKETITETTFYVDYDDL 269
DQ 309 NQLRQODEDFRKQLEBKGRKTAENVMTELT-----MEINKWRLLYEEL 353

RESULT 19
US-10-257-377-2
; Sequence 2, Application US/10257377
; Publication No. US20040037834A1
; GENERAL INFORMATION:
; APPLICANT: B. Michael R. Woloski,
; APPLICANT: Ashley M. Williams,
; APPLICANT: Terrance J. Sereda,
; APPLICANT: Deanna J. Wiebe
; TITLE OF INVENTION: RHAMM Peptide Conjugates
; FILE REFERENCE: BERE-1001US
; CURRENT APPLICATION NUMBER: US/10/257,377
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/198,613
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 794
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-257-377-2

Query Match          7.4%; Score 111; DB 15; Length 794;
Best Local Similarity 20.5%; Pred. No. 2.9;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVPMQTFDETTELKELSRFKQVYSQAASV 60
DQ 416 LDNLLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 472
QY 61 LVGDIKTLLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLDFEYNEKKAQAQKDLIKVL 120
DQ 473 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 522
QY 121 DDGITKLNKAQKSLI-VSSQSPFNASGKLAL-DSQLTN-DPSEKSSYFQSQVDKIRKEA 177
DQ 523 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLNLTLQEKVMAEKSVEDVQOOI 582
QY 178 YAGAAAGVAVGPFGLIISYIAAGVVEGKLIPELKNK-----LKSVQNF-----TTLS 226
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-978-309A-77

Query Match          7.4%; Score 111; DB 10; Length 476;
Best Local Similarity 20.5%; Pred. No. 1.5;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVPMQTFDETTELKELSRFKQVYSQAASV 60
DQ 98 LDNLLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 154
QY 61 LVGDIKTLLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLDFEYNEKKAQAQKDLIKVL 120
DQ 155 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 204
QY 121 DDGITKLNKAQKSLI-VSSQSPFNASGKLAL-DSQLTN-DPSEKSSYFQSQVDKIRKEA 177
DQ 205 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLNLTLQEKVMAEKSVEDVQOOI 264
QY 178 YAGAAAGVAVGPFGLIISYIAAGVVEGKLIPELKNK-----LKSVQNF-----TTLS 226
DQ 265 LTAESTNQ-----EVA-----RMVQDLQNRSTLKEEIKETITSSFLEKITDLK 307
QY 227 NTVKQANKDI-----DAAKLKLTEIAAIGEIKETITETTFYVDYDDL 269
DQ 308 NQLRQODEDFRKQLEBKGRKTAENVMTELT-----MEINKWRLLYEEL 352

RESULT 19
US-10-257-377-2
; Sequence 2, Application US/10257377
; Publication No. US20040037834A1
; GENERAL INFORMATION:
; APPLICANT: B. Michael R. Woloski,
; APPLICANT: Ashley M. Williams,
; APPLICANT: Terrance J. Sereda,
; APPLICANT: Deanna J. Wiebe
; TITLE OF INVENTION: RHAMM Peptide Conjugates
; FILE REFERENCE: BERE-1001US
; CURRENT APPLICATION NUMBER: US/10/257,377
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/198,613
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 794
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-257-377-2

Query Match          7.4%; Score 111; DB 15; Length 794;
Best Local Similarity 20.5%; Pred. No. 2.9;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVPMQTFDETTELKELSRFKQVYSQAASV 60
DQ 416 LDNLLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 472
QY 61 LVGDIKTLLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLDFEYNEKKAQAQKDLIKVL 120
DQ 473 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 522
QY 121 DDGITKLNKAQKSLI-VSSQSPFNASGKLAL-DSQLTN-DPSEKSSYFQSQVDKIRKEA 177
DQ 523 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLNLTLQEKVMAEKSVEDVQOOI 582
QY 178 YAGAAAGVAVGPFGLIISYIAAGVVEGKLIPELKNK-----LKSVQNF-----TTLS 226
```



Db 583 LTBSTNQ-----EYA-----RMVQDLQNRSTLKEBEIKETSFLEKITDLK 625

Qy 227 NTVKQANKDI-----DAKLUKLTTEIAAIGEIKTETETTRFVVDYDDL 269

Db 626 NQLRQODEDFRQLEKGRKRTAKENVTMLT-----MEINKWRLLYEEL 670

RESULT 20

US-10-282-122A-71606

; Sequence 71606, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Heselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 71606

; LENGTH: 938

; TYPE: PRT

; ORGANISM: Staphylococcus haemolyticus

US-10-282-122A-71606

Query Match 7.4%; Score 111; DB 15; Length 938;

Best Local Similarity 18.6%; Pred. No. 3.6;

Matches 60; Conservative 64; Mismatches 110; Indels 88; Gaps 11;

Qy 9 TVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFQKQYSAQSVLVGDIKTL 68

Db 189 TINKIKNAVYAADSLPQINKIADRI---EYLDHDDLDKYANQFR-----ALGNYKGD 240

Qy 69 LMDSDQKFEATQVYEWCGVATQLLAAYILLFDEY-----NEKKAQAK--- 113

Db 241 ILDAQQLNDVNAI-----PSLNEKAKLILALNEYPNTEKLDVANSIDIPQFPKINR 295

Qy 114 --DILIKVLDGDTIKLNEAQKSLVSSQ---SFFNASGKLLALDSQITNDFSEKSS--YF 166

Db 296 GVDIASEGFDLANRLNDAQGYLTSAQQRVDYQEAAGRAQYEVNQAANSALRQOSTSGLP 355

Qy 167 OSQVDKIRKEAYAGAAAGVAGPFGLLIISYIAAGVVEGKLIPELKNKLSQVNFPTLS 226

Db 356 QYQIQKLSYD-----NSQDTVNDNQIVSN 379

Qy 227 NTVKQANKDIDAAKLUKLT-----TEIAAIGEIKTETETTRFVVDYDDLMLSL-- 273

Db 380 NDVKSMNSALAEALLTLSSNDNQAKATQSDIKALADISYGVIGSNRPTFENDMLRLNKT 439

Qy 274 -LKEAAK---KMINTCNEYQKR 291

Db 440 RLENSSKSNQQLIDVLKLEKR 461

RESULT 21

US-09-884-696-5

; Sequence 5, Application US/09884696

; Publication No. US20030035809A1

; GENERAL INFORMATION:

; APPLICANT: GEORGE, LISLE W

; APPLICANT: ANGELOS, JOHN A

; APPLICANT: HESS, JOHN F

; TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES

; TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA

; TITLE OF INVENTION: BOVIS INFECTIONS

; FILE REFERENCE: 481.06

; CURRENT APPLICATION NUMBER: US/09/884,696

; CURRENT FILING DATE: 2001-06-19

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5

; LENGTH: 1023

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-884-696-5

Query Match 7.4%; Score 111; DB 10; Length 1023;

Best Local Similarity 21.3%; Pred. No. 4.1;

Matches 71; Conservative 56; Mismatches 116; Indels 90; Gaps 11;

Qy 2 TEIVADKTVEVVK-----NAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFQKE 53

Db 155 TALSSMKIDELIKKQKSGNVSSSELAKASIELINQLVDTA-----ASLNNVNSFSQ 207

Qy 54 YSOAASVL-----VGIKTLMDSDQKQYFATQVYEWCGVATQLLAAYILL--FDEY 104

Db 208 LNKLGVLSTNKLNGVGN--KLQNLPLNDINIGAGLDTV---SGILSAISAFILSNADAD 263

Qy 105 NEKKAQAKDILIKVLDD---GITKLEAQKSLVSSQSFNNASGKLLALDSQITNDFSE 161

Db 264 TGTAAAGVELTTKVLGNVGKISQYIIAQAAGLST----- 302

Qy 162 KSSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIIS---YSIAAGVVEGKLIPELKNKLS 217

Db 303 -----AAAAGLIASVVTLSPLSFLSIADKFRANKIEEVSQRFKK 344

Qy 218 VQNFPTLSNTVKQANKDIDAAKLUKLTTEIAAIGEIKTETETTRF-----YVDYDDL 270

Db 345 LGYDGDLSLAAFHKETGAIDASLTRISTVLASVSSGISAATTSLVGAPVSVLVGAVTGI 404

Qy 271 LSLKKAQK-----KMINTCNEYQKRHGK 294

Db 405 ISGILEASKQAMPEHVASKMADVIAEWKRGK 437

RESULT 22

US-10-032-585-7519

; Sequence 7519, Application US/10032585

; Publication No. US20030180953A1

; GENERAL INFORMATION:

; APPLICANT: Terry, Rosmer D.

; APPLICANT: Bo, Jiang

; APPLICANT: Charles, Boone

; APPLICANT: Howard, Bussey

;; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
;; FILE REFERENCE: 10182-005-999  
;; CURRENT APPLICATION NUMBER: US/10/032,585  
;; CURRENT FILING DATE: 2001-12-20  
;; NUMBER OF SEQ ID NOS: 8000  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 7519  
;; LENGTH: 1171  
;; TYPE: PRT  
;; ORGANISM: Candida albicans  
US-10-032-585-7519

Query Match 7.4%; Score 111; DB 14; Length 1171;  
Best Local Similarity 18.3%; Pred. No. 4.9;  
Matches 40; Conservative 53; Mismatches 87; Indels 38; Gaps 5;

QY 103 EYNEKASAKQDILIK-VLDGGITKLENAQKSLVSSQFNASGKLLALDSQLTWDFS 160  
DB 289 EENLTGDSMELEIKENQNSDRLNLTARDIANDLNTTEKNKHTKLIQLSQIKOOLA 348  
QY 161 EKSSVFQSQVDKIR-----KEAVAGAAAGVAGVPGFGLIISYSIAAGVVEGKLIPEL 211  
DB 349 SNOQTFDQENYKQSQNELVQKKEEYANKQELLSTLSTGISTGNVTSGVI--TQLDNV 406  
QY 212 KNLKSVQNFPTLNTVQKQKDAKLLTTEIAAIGELKTETETTRFYVD-----265  
DB 407 KSKLASNFIKTSTKLNHLQOQMESQMKLT-----KAKTENETLLSAIELHRQFI 459  
QY 266 -----YDMLSLKKEAAKMMTNCNEYOKR 291  
DB 460 KTKQLEIDSCLGFEPTKIHQLRDQESLISQNKFNQ 497

RESULT 23  
US-10-282-122A-51864  
; Sequence 51864, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A US/10/282,122A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308

;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.

;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 51864  
;; LENGTH: 1163  
;; TYPE: PRT  
;; ORGANISM: Clostridium acetobutylicum  
US-10-282-122A-51864

Query Match 7.3%; Score 110; DB 15; Length 1163;  
Best Local Similarity 17.9%; Pred. No. 5.9; Indels 152; Gaps 14;  
Matches 73; Conservative 69; Mismatches 114;

QY 1 MTEIVADKTVEVKNATETADG-----ALDLYNKYLDQVIPWOTFDETIK-----45  
DB 109 IVDITGDB-VEVLEGAKSVNEKQEIIGLSLDDFTRTV--VLPQGRFSBFLKLEGKERR 165  
QY 46 -----ELSRFKQEYSQAASVLVGDIK-----TLMDSQ 73  
DB 166 NWLERLNLQYDELSFKLARKIRKEREKENVLVGLKGYENINEDVLKERRELLKENN 225  
QY 74 DKYFEATQTVVEMCGVATQLLAAVILLDFEYN-----EKKASAKQDILIKV-- 119  
DB 226 DFFNEASK-----EYLKABEYNEGKEVWGLQIEIEEKURVRKDLMEKKDE 271  
QY 120 --LDGGITKLENAQKSLVSSQFNASGKLLALDSQLTN-----DSEKSS 164  
DB 272 IDLKEKRALGESSKVKPYIDNVENTLKQIDILKEQILSRENTWKAISLEKDEMEKKLS 331  
QY 165 YFQSQVDKIRKEAYAGAAAGVAGVPGFGLIISYSIAAGVVEGKLIPELKNKLSVQNFPT 224  
DB 332 IAKDNKEK-----ALPFWIKHHIILDAIKERKOLLNLIKLEKRLQKIEK 377  
QY 225 LS-----NTVQANKOIDAAKLKTTEIAAIGIKTETE-----258  
DB 378 LSLEASNKEELIKQNIKDLSLTQNLKESKIDNLKVPPEYKKNKINEGIFLLRNYDEKL 437  
QY 259 -----TTRFYVDYD-----DLMLSLKKEAAKMMTNCNEYOKR 291  
DB 438 KHRNKLGLDCDFQVDFEKAQKSKKEMLFNKLSEERSKL-----DITYKK 481

RESULT 24  
US-10-425-114-60225  
; Sequence 60225, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 60225  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3607-037-H5\_FLI pep  
US-10-425-114-60225

Query Match 7.3%; Score 109.5; DB 15; Length 420;  
Best Local Similarity 19.1%; Pred. No. 1.6; Indels 69; Gaps 16;  
Matches 60; Conservative 82; Mismatches 103;

QY 5 VADKTVEVKNATETAD--GALDLYNKYLDQVIPWOTFDETIKELSRFKQEYSQAASVL 61

Db 14 ISDRIAELEDEQARQQIHALEVRKMHQ-----CEKLESKQVNEKEDLVAAILAS 69  
Qy 62 VGDIKTLMD--SODKYFEATQTVY-----EWCGVATQLLAAAYILLFDEYNEKKASAKD 114  
Db 70 KNEVEDLKGDMVSAAKHFA-QLVHRDHEIEKCKQEAQVS-----EKYFHEKSTLESE 122  
Qy 115 I-----LIKVLDDGIIKLN-----BAQ-KSLIVSSQSFNNASKLLA-----LDSQL 155  
Db 123 IORIQEVVKFEENLTQVADEKLEAQLEQTSNDLDDSSAEIILKQEIINKDQERL 162  
Qy 156 TNDPEKSSYFQ--SQVDKIRKEAYAGAAGVAGPFGLIISYSIAAGVVEGKLIPELKN 213  
Db 183 ENDSNEKSVLEERAMELEQVRQLEDSEASME-----LQTTIKNLQVLOEK--AELEN 236  
Qy 214 KLSVQNVFTLSTNTVKQANIDAAKLUETTEIAAIGEIKTETETTRFFYVDVDDMLSL 273  
Db 237 RMKDAE-----QATSDLSNLVSLDGKKA-----TEAQLQOLHAERAEATLE- 279  
Qy 274 LKZAAKKNWINTCNE 287,  
Db 280 ----SEKQVSELNQ 289

## RESULT 25

US-09-841-260-139  
; Sequence 139, Application US/09841260  
; Publication No. US20030175700A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Probst, Peter  
; APPLICANT: Stromberg, Erika Jean  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS  
; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.515  
; CURRENT APPLICATION NUMBER: US/09/841.260  
; CURRENT FILING DATE: 2001-04-23  
; NUMBER OF SEQ ID NOS: 140  
; SEQ ID NO 139  
; LENGTH: 660  
; TYPE: PRT  
; ORGANISM: Chlamydia trachomatis  
US-09-841-260-139

Query Match 7.3%; Score 109.5; DB 10; Length 660;  
Best Local Similarity 20.6%; Pred. No. 3;  
Matches 65; Conservative 56; Mismatches 111; Indels 83; Gaps 15;

Qy 10 VEVVKNAIETADGALDLYNKYL-----DQV-----IPWQTFDE 42  
Db 157 VNNIKKALEAQKOTIDKLNKLVTLQNKSLTEVLKTTDSADQIPAINSOLEINKNSADQ 216  
Qy 43 TIKELSRFKQEYSQAASVVLGDIKTLMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFD 102  
Db 217 IIKDLERQNTSY-EAVLTNAGEVIKASSEAGIKLGQALQSIDV---AGDQSQAAVLQAAQ 272  
Qy 103 EYNEKKAQKODILIKVLDGDIKTLNEAQKSLIVSSQSFNNASKLLALDSQLTNDPSEK 162  
Db 273 NNSPDNIAATKEL-----IDAAETKVNLKQ-----EHTG---LTDSPLVKKAEEQ 315  
Qy 163 SSVFQSQVDKIRKEAYAGAAGVAGPFGLIISYSIAAGVVEGK-----LIPELKNK 214  
Db 316 ISQAQKDIQEIKP---SGSDIPIV-GPSSGAASAGSAGALKSSNNSGRISLLDDVDNE 371  
Qy 215 LKSV--QNFFTLSTNTVKQAN-----KIDAAKLKLTTE-----IAAIGEIKTETETTR 261  
Db 372 MAALALQGF-----RSMIEQFNVNPNATAKELQAMEAQLTAMSDQLVGADGELPAEIQAIK 427  
Qy 262 FYVDYDDLMLSLKE 276  
Db 428 -----DALAQALQ 436

## RESULT 26

US-10-007-693-139  
; Sequence 139, Application US/10007693  
; Publication No. US20020146776A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Probst, Peter  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT  
; TITLE OF INVENTION: AND DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.515C2  
; CURRENT APPLICATION NUMBER: US/10/007.693  
; CURRENT FILING DATE: 2001-12-05  
; NUMBER OF SEQ ID NOS: 157  
; SEQ ID NO 139  
; LENGTH: 660  
; TYPE: PRT  
; ORGANISM: Chlamydia trachomatis  
US-10-007-693-139

Query Match 7.3%; Score 109.5; DB 13; Length 660;  
Best Local Similarity 20.6%; Pred. No. 3;  
Matches 65; Conservative 56; Mismatches 111; Indels 83; Gaps 15;

Qy 10 VEVVKNAIETADGALDLYNKYL-----DQV-----IPWQTFDE 42  
Db 157 VNNIKKALEAQKOTIDKLNKLVTLQNKSLTEVLKTTDSADQIPAINSOLEINKNSADQ 216  
Qy 43 TIKELSRFKQEYSQAASVVLGDIKTLMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFD 102  
Db 217 IIKDLERQNTSY-EAVLTNAGEVIKASSEAGIKLGQALQSIDV---AGDQSQAAVLQAAQ 272  
Qy 103 EYNEKKAQKODILIKVLDGDIKTLNEAQKSLIVSSQSFNNASKLLALDSQLTNDPSEK 162  
Db 273 NNSPDNIAATKEL-----IDAAETKVNLKQ-----EHTG---LTDSPLVKKAEEQ 315  
Qy 163 SSVFQSQVDKIRKEAYAGAAGVAGPFGLIISYSIAAGVVEGK-----LIPELKNK 214  
Db 316 ISQAQKDIQEIKP---SGSDIPIV-GPSSGAASAGSAGALKSSNNSGRISLLDDVDNE 371  
Qy 215 LKSV--QNFFTLSTNTVKQAN-----KIDAAKLKLTTE-----IAAIGEIKTETETTR 261  
Db 372 MAALALQGF-----RSMIEQFNVNPNATAKELQAMEAQLTAMSDQLVGADGELPAEIQAIK 427  
Qy 262 FYVDYDDLMLSLKE 276  
Db 428 -----DALAQALQ 436

## RESULT 27

US-10-762-058-139  
; Sequence 139, Application US/10762058  
; Publication No. US20040137007A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Probst, Peter  
; APPLICANT: Stromberg, Erika Jean  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS  
; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.515D1  
; CURRENT APPLICATION NUMBER: US/10/762.058  
; CURRENT FILING DATE: 2004-01-15  
; NUMBER OF SEQ ID NOS: 140  
; SEQ ID NO 139  
; LENGTH: 660  
; TYPE: PRT  
; ORGANISM: Chlamydia trachomatis  
US-10-762-058-139

Query Match 7.3%; Score 109.5; DB 16; Length 660;  
Best Local Similarity 20.6%; Pred. No. 3;  
Matches 65; Conservative 56; Mismatches 111; Indels 83; Gaps 15;

Qy 10 VEVVKNAIETADGALDLYNKYL-----DQV-----IPWQTFDE 42

Wed Jan 5 14:16:57 2005

```

Db      157 VNNIKKALEAKQOTIDKLNKLVTLQNKQSLTEVLKTDSDAQIPAINSOLEINKNSADQ 216
Qy      43  TIKELSRKQKQYQAAVILVGDITKLLMDSQDKYFEATQVYEWCGVATQLLAAVILLFD 102
Db      217 IIKDLERQNTISY-EAVLTNAGEVIKASSEAGIKGLQALQSIVD---AGDSQAVLQAOQ 272
Qy      103 EYNEKASAKDILIKVLDDGIVKLNKAEQKSLIVSSQSFNNASGKLLALDSQTNDFSEK 162
Db      273 NNSPDNIAATKEL-----IDAEIKVNLKQ-----EHTG---LTDSPLVKKAEEQ 315
Qy      163 SSYFSQSDVKIRKAYAGAAVGVAGPGLIISVIAAGVVEGK-----LIPKLNK 214
Db      316 ISQAQKQIOEIKP---SGSDIPIV-GPSSGAASAGSAGALKSSNNSGRISLLDLDVDNE 371
Qy      215 LKSV--QNFFTLNTVTKQAN-----KQIDAAKKLKLTTE---IAAIGEIKTETETTR 261
Db      372 MAATIALQGF---RSMIEQFVNNPNATAKELQAMEAQUTAMSDQLVGADGELPAEIQAIK 427
Qy      262 FYVDYDDLMLSLKE 276
Db      428 -----DALAQALKQ 436

RESULT 28
US-10-369-493-1061
; Sequence 1061, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1061
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Methanococcus jannaschii
US-10-369-493-1061

Query Match      7.2%; Score 109; DB 14; Length 1005;
Best Local Similarity 19.1%; Pred. No. 5.8;
Matches 58; Conservative 55; Mismatches 93; Indels 98; Gaps 11;

Qy      7  DKTVEVVVKNIAETADGALDLYNKYLQVTPWQTFDE---TIKELSRF-----KQYEQ 56
Db      605 DEILEDIKSQLNFKF---NFYQYLSAVSYLNSVDEGINRRIKEINIVSGWNEKCKRE 661
Qy      57 AASVL-----VGDITKLLMDSQDKYFEATQVYEWCGVATQLLAAVILLFDEYNEKKA 110
Db      662 ELNKLREDEIREINLKDKNELKKEK-----LIEIENRRSLKFKYKEYLGL 710
Qy      111 AQKDILIKVLDDGIVKLNKAEQKSLIVSSQSFNNASGKLLALD----- 152
Db      711 TEKLEELKNIKDGLLEEI-----INTCNSKILADINIKKYNKEDIYVYLNK 757
Qy      153 -----SQTNDPSEKSSYFQSOVDKIRKEAVAGAAVGVAGPGLIISVIAAGVVEGKI 208
Db      758 ILEVNKELNDEERISYNQKLE-----INYN-----EEBHKI 792
Qy      209 FEL-----KNKLSVQNFFTLNTVTKQANQKIDIAKKLKLTTEAAIGEIKTETTRFYV 264
Db      793 KELVENKQELNDVREQKTEIETGIEYLLKQVESLKARL-----KEMENLEKEKEKLTQV 848
Qy      265 DYDD 268

```

```

Db      849 EYLD 852

RESULT 29
US-09-978-309A-79
; Sequence 79, Application US/09978309A
; Publication No. US20030100490A1
; GENERAL INFORMATION:
; APPLICANT: Cruz, Tony
; APPLICANT: Pastrak, Aleksandra
; APPLICANT: Turley, Eva A.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to
; TITLE OF INVENTION: Injury and Other Proliferating Cell Disorders Regulated by
; TITLE OF INVENTION: Hyaladherin and Hyaluronans
; FILE REFERENCE: 033352-010
; CURRENT APPLICATION NUMBER: US/09/978,309A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 09/685,010
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 09/541,522
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/127,457
; PRIOR FILING DATE: 1999-04-01
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-978-309A-79

Query Match      7.2%; Score 108.5; DB 10; Length 476;
Best Local Similarity 20.2%; Pred. No. 2.3;
Matches 59; Conservative 63; Mismatches 109; Indels 61; Gaps 15;

Qy      1  MTEIVADKTEVVVKNIAETADGALDLYNKYLQVTPWQTFDETIKELSRKQYEQAAV 60
Db      99 LDNLREKEVELEKHAHAQAAILTAQEKYNTA---QSLRDVTAQLESQVQEKYNTAQS 155
Qy      61 LVGDIKTLMDSQDKYFEATQVYEWCGVATQLLAAVILLFDEYNEKKAQKIDILIKVL 120
Db      156 L-RDVTAAQLESQEKYNTAQSRLD---VTAQLESEQ---EKYND-TAQSRLDVTQAQ-L 205
Qy      121 DDGITKLNKAEQKSLIVSSQSFNNASGKLLAL-DSOLTN-DFSEKSSYFQSOVDKIRKEA 177
Db      206 ESQVQEKYNTAQSRLDVTAAQLESQYKSTLKEIEDLKENLTLQEKVMAEKSVEDVQOQI 265
Qy      178 YAGAAAGVAGPGLIISVIAAGVVEGKLIPELKNK-----LKSQNF-----FTLSN 227
Db      266 LTAESTNQ-----EYA-----RMVQDLQNRSLKKEBEIKETSFLKIDTLKN 308
Qy      228 TVQKANKDI-----DAAKKLKLTTEAAIGEIKTETTRFYVDYDDL 269
Db      309 QLROQDEDFRQLEEKGRKTAENNVNTELT-----MEINKWLLLYEEL 352

RESULT 30
US-10-739-930-6691
; Sequence 6691, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 6691
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana

```



```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazov, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 107270
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_11638C.1.pep
US-10-437-963-107270

Query Match          7.2%; Score 108; DB 16; Length 737;
Best Local Similarity 22.0%; Pred. No. 4.6;
Matches 71; Conservative 63; Mismatches 126; Indels 62; Gaps 14;

QY 4 IVADKTVEVVKNAIETAD---GALDLYNKYLDQVVPWQTFDETTELKLSRFKQESYQ--- 56
DB 249 LVAQKQKNCIBAEIERLKMELGALTEANAAKAFDTQN--EITKELEDLTKLEEIKTN 307
QY 57 --AASVLVDGIKTLMSQDKYFRATQTVYEWCGVATQLLAAYILLDFEYNEKKASAKQD 114
DB 308 KDLAESNGKLSRELLSAEEKYSQAEV-----KYLQVMGA--VVEAKAAKAFAAEKE 362
QY 115 ILIKVLDGIIKLNAAOKSLVSSQFNNSGKLLALDSLTNDPSEKSSYFQSQVDKIR 174
DB 363 DIMKESDNLKRVKXIEIQSKLLVSENDELSEILSM-----KQKHGFVEVITSLK 414
QY 175 KEAVAGAAGVVGPPFGLIISYSIAAGVVEG---KLIPELKNLKSQVN----- 220
DB 415 KELGALEAKEI-----TTKAPVEVTEILKELEDLKRKVEIQTNKDLVEVGND 464
QY 221 --FTTLNTVTKQANKOIDAQKLKLTETIAAIGIKETETETTRFVVDYDDLMLSL--LKEA 277
DB 465 KRLDLVLSAQKQSIILEVEANLKM--ELGALVEAK-EVATKAF--DAEKAKITKELEDV 519
QY 278 AKKMINTCNEVKRHKGTLP 299
DB 520 KKKM-----EEIQVKKDLVE 534

RESULT 34
US-10-205-647A-6
; Sequence 6, Application US/10205647A
; Publication No. US20040010812A1
; GENERAL INFORMATION:
; APPLICANT: University of Manitoba
; APPLICANT: Manitoba Cancer Treatment and Research Foundation
; APPLICANT: TURLEY, Eva A.
; APPLICANT: ENTWISTLE, Joycelyn
; TITLE OF INVENTION: HUMAN HYALURONAN RECEPTOR
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; STREET: 181 Freedman Crescent, Room 361
; CITY: Winnipeg
; STATE: Manitoba
; COUNTRY: Canada
; ZIP: R3T 5V4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/10/205,647A
; FILING DATE: 23-JULY-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CA97/00240
; FILING DATE: 10-APR-1996
; APPLICATION NUMBER: GB 9607441.4
; FILING DATE: 10
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-10-205-647A-6

Query Match          7.1%; Score 107.5; DB 15; Length 630;
Best Local Similarity 19.8%; Pred. No. 4.1;
Matches 58; Conservative 65; Mismatches 107; Indels 63; Gaps 15;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVVPWQTFDETTELKLSRFKQESQAASV 60
DB 253 LONLLREKEVELEKHTAAHAQAIIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 309
QY 61 LVGDITKILMSQDKYFRATQTVYEWCGVATQLLAAYILLDFEYNEKKASAKQDILIKVL 120
DB 310 L-RDVTQAQLESEQEKYNDTAQSLRD---VTAQLESEQ-----EKYND--TAQSLRDVTAQL- 359
QY 121 DGGITKLNAAQSKLL--VSSQFNNSGKLLAL--DSQLTN--DFSEKSSYFQSQVDKIRKEA 177
DB 360 -ESQEKYNDTAQSLRDVTAQLESYKSTLKEIEDLKLENTLQEKVMAEKSVEDVQOQI 418
QY 178 YAGAAAGVVGPPFGLIISYSIAAGVVEGKLIPELKNK-----LKSQVNF-----TTLS 226
DB 419 LTAESTNQ-----EVA-----RMVODLQNRSLKEEIKETITSSFLEKITDLK 461
QY 227 NTVKQANKOI-----DAAKLTJTTIAAIGIKETETETTRFVVDYDDL 269
DB 462 NQLRQDQDEPRKOLEKGRKTAENVMTELT-----MEINKWLLIYEEL 506

RESULT 35
US-09-978-309A-48
; Sequence 48, Application US/09978309A
; Publication No. US20030100490A1
; GENERAL INFORMATION:
; APPLICANT: Cruz, Tony
; APPLICANT: Pastrak, Aleksandra
; APPLICANT: Turley, Eva A.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to
; TITLE OF INVENTION: Injury and Other Proliferating Cell Disorders Regulated by
; TITLE OF INVENTION: Hyaladerin and Hyaluronans
; FILE REFERENCE: 033352-010
; CURRENT APPLICATION NUMBER: US/09/978,309A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 09/685,010
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 09/541,522
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/127,457
; PRIOR FILING DATE: 1999-04-01
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 631
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-978-309A-48

Query Match          7.1%; Score 107.5; DB 10; Length 631;
Best Local Similarity 19.8%; Pred. No. 4.1;
Matches 58; Conservative 65; Mismatches 107; Indels 63; Gaps 15;

```







RESULT 40  
US-10-028-248A-105  
; Sequence 105, Application US/10028248A  
; Publication No. US20030235882A1  
; GENERAL INFORMATION:  
; APPLICANT: Shinkets, Richard  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Vernet, Corine  
; APPLICANT: Casman, Stacie  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Shenoy, Sureesh  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Gangolli, Esha  
; APPLICANT: Miller, Charles  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Li, Li  
; APPLICANT: Taupier Jr, Raymond J  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Colman, Steven  
; APPLICANT: Tchernev, Velizar  
; APPLICANT: Si, Jingsheng  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Stone, David  
; APPLICANT: Sciore, Paul  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Rothenberg, Mark  
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods  
; FILE REFERENCE: 21402-222  
; CURRENT APPLICATION NUMBER: US/10/028,248A  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/256619  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/262959  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: 60/272408  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/285189  
; PRIOR FILING DATE: 2001-04-20  
; PRIOR APPLICATION NUMBER: 60/308039  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: 60/311266  
; PRIOR FILING DATE: 2001-08-09  
; NUMBER OF SEQ ID NOS: 211  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 105  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-10-028-248A-105  
Query Match 7.0%; Score 106; DB 14; Length 1961;  
Best Local Similarity 20.4%; Pred. No. 25;  
Matches 69; Conservative 69; Mismatches 120; Indels 80; Gaps 16;  
Qy 3 EIVADKTVVVKNAIETADGALDLYNKYLDQV-IPWQTFDETIK-----ELSRFKQEVSOA 57  
Db 1140 ELEALKT--ELEDTLDTAAQQLRSRQEVSVILKKTLEDEAKTHPAQIQEMRQKHSQA 1197  
Qy 58 ASVLVGDKITLLMDSQDKYFEAT-----QTVYECG-VATOLLAAYILL----FDEYNEK 107  
Db 1198 VEELAEQLE-----QTKRVKATLEKAKQTLNERGELANEVKA---LLQKGDSEHKRK 1248  
Qy 108 KASAQ-KDILIKV-----LDGKITKLENAQKSLVSSQSFNNASGKIL-----ALDS 153  
Db 1249 KVEAQLQELQVKFGEGERVRLTELADKYSKLQVELDSTVGLLNQSDSKSKLTKDFSALES 1308  
Qy 154 QLTN-----DFSEKSSYFQSQVDKIRKEAYAGAAAGVAGPGLI 193

Db 1309 QLODTQELQENRQKLSLSTKLQWEDENKSNFRELEEEERAKRN----- 1355  
Qy 194 ISYSIAAGVVEGKLIPELKNKJKSVQNFPTTSLNTVQKANKDIDAAKLKTTEIAAIGBI 253  
Db 1356 LEKQIATLHAQ---VTDMKKMEDGVGCLETAEEAKRRLQKLEGLSQRLEEKVAAYD- 1410  
Qy 254 KTETETRTFYVDYDDLMLSLKKEAAKMNITCNEYQKR 291  
Db 1411 KLEKTKTRLQOELDDLVL--DHQRQSVSNLEKKQKK 1446  
RESULT 41  
US-10-107-782-105  
; Sequence 105, Application US/10107782  
; Publication No. US20040018970A1  
; GENERAL INFORMATION:  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Casman, Stacie  
; APPLICANT: Colman, Steve  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Gangolli, Esha  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Li, Li  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Miller, Charles  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Rothenberg, Mark  
; APPLICANT: Sciore, Paul  
; APPLICANT: Shenoy, Sureesh  
; APPLICANT: Shinkets, Richard  
; APPLICANT: Si, Jingsheng  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Stone, David  
; APPLICANT: Taupier, Raymond, Jr.,  
; APPLICANT: Tchernev, Velizar,  
; APPLICANT: Vernet, Corine  
; APPLICANT: Zerhusen, Brian  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF  
; FILE REFERENCE: 21402-222CIP  
; CURRENT APPLICATION NUMBER: US/10/107,782  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 10/028,248  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/256,619  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/262,959  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: 60/272,408  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/285,189  
; PRIOR FILING DATE: 2001-04-20  
; PRIOR APPLICATION NUMBER: 60/308,039  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: 60/311,266  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 60/279,344  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 215  
; SOFTWARE: Curaseqlist version 0.1  
; SEQ ID NO 105  
; LENGTH: 1961  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-107-782-105  
Query Match 7.0%; Score 106; DB 15; Length 1961;  
Best Local Similarity 20.4%; Pred. No. 25;  
Matches 69; Conservative 69; Mismatches 120; Indels 80; Gaps 16;

QY 3 EIVADKTVEVVKNAIETADGALDLYNKYLDQV-IPWOTDETIK-----ELSRFKQESOA 57  
 Db 1140 ELEAKT--ELEDLTDTAAQQLRSKREGEVSILKKTLEDEAKTHEAQIQEMRQKHSOA 1197  
 QY 58 ASVLVGDITKLLMDSQKYEAT-----QTVYEWCG-VATOLLAAYILL----FDEYNEK 107  
 Db 1198 VEELAEQLE-----QTRVKATLEKAKQTLNERGELANEVKA---LLOGKGDSEHK 1248  
 QY 108 KASQAQ-KDILIKV-----LDDGITKLENAQKSLVSSQSFNNASGKLL-----ALDS 153  
 Db 1249 KVEAQLQELQVKFSEGERVTELADKVS KLQVELDSVTGLNQSDSKSKLTKDPSALES 1308  
 QY 154 QLTN-----DPSEKSYQSQVDKTKRKEAYAGAAAGVAGFPGLI 193  
 Db 1309 QLODTQELLQENRQKLSLSTKLQMEDEKNSFREQLSEEEAKRN-----1355  
 QY 194 ISYSTAAGVVEGKLIPELKNLKSQVQFFTTLSNTVQANKDIDAAKLKLTTTIAAIGEI 253  
 Db 1356 LEKQATLHAQ---VTDMKKKMGVGCLETAEEAKRRLQKDLGLSLQRLEEKVAAYD-- 1410  
 QY 254 KTETETTRFYVDYDMLSLKEAKKMINTCNBYQKR 291  
 Db 1411 KLEKTKRLQELDLDDLVL--DHQOSVSNLEKKQK 1446

RESULT 42  
 US-09-815-242-5815  
 ; Sequence 5815, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Cart, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 5815  
 ; LENGTH: 2076  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus aureus  
 US-09-815-242-5815  
 Query Match 7.0%; Score 106; DB 9; Length 2076;  
 Best Local Similarity 22.4%; Pred. No. 27;  
 Matches 72; Conservative 56; Mismatches 126; Indels 68; Caps 11;

QY 10 VEVVKNAIETA-----DGALDLYNKYLDQVIPWOTDETIKELSRFKQESOAASVLVGD 65  
 Db 1674 LDAIRNTLDTQDERDVAIDTLNKIVNTIKNDIAQNKTAEBVORTDGDND-----NI 1726

QY 66 KTLMDSQDKYFEATQTVYEWCGVATOLLAAYILLFDEYNEKASQAQKDLIKVLDDGIT 125  
 Db 1727 KVILPKVQVP-AARQSV-----GVKAEQAQNALIDQSDLSSTEERLAAKHLVEQALNOAID 1781  
 QY 126 KLNBAQKSLVSSQSFN--NASGK-----LLALDSQLTNDPS-----EK 162  
 Db 1782 QINHADKTAQVNODSINAQNIISKIKPATTVKATALQIQIATNKINLIKANNEATDEE 1841  
 QY 163 SSYQSQVDKTKRKEAYAGAAAGVAGFPGLIISYSIAAGVVEGKLIPELKNLKSQVQFF 222  
 Db 1842 QNIAIAQVEKELIKAKQIASAVTNAD---VAYLLHDEKNEIREIEBPVNRKASAREQL 1897  
 QY 223 TSLNTVQANKDIDAAKLKLTTT-----IAAIGEIKTB-----TETTR 261  
 Db 1898 TLFNFKKQAIE---ANIQTVEERNSILAQLONIYDTALQIQDQDRSNAQVDTKASLN 1953  
 QY 262 FVVDYDMLSLKEAKKMIN 283  
 Db 1954 LOTIHLDLVHPFKPDDEKTN 1975

RESULT 43  
 US-09-815-242-5639  
 ; Sequence 5639, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Cart, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 5639  
 ; LENGTH: 2086  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus aureus  
 US-09-815-242-5639

Query Match 7.0%; Score 106; DB 9; Length 2086;  
 Best Local Similarity 21.9%; Pred. No. 28;  
 Matches 75; Conservative 37; Mismatches 115; Indels 116; Gaps 17;

QY 9 TVEVVKNAIETADGALDLYNKYLDQVI-----PWOTDETIKELSRFKQESOAASVLV 62  
 Db 264 TITQYNNALHNAAQQQINTAKTAEQAQVNNERATPOQVSDALTK-----VRAAQTKI 314  
 QY 63 GDIKTLMDSDQ-----KYFEATQTVYEWCGVATOLLAAYILLFDEYNEKASQAQK 113  
 Db 315 NEAKALLQNKEDNSQVTSKNLQSSVQNPVSTTGMTCQSI-----DYNNAKKEAET 367



```

; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,849
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53523
; LENGTH: 742
; TYPE: PR1
; ORGANISM: Clostridium difficile
US-10-282-122A-53523

Query Match 7.0%; Score 105.5; DB 15; Length 742;
Best Local Similarity 23.5%; Pred. No. 7.5; Indels 69; Gaps 14;
Matches 72; Conservative 45; Mismatches 120;

QY 18 ETADGALDLNKKYLDVPIPWQTEDETIKELSRFKQBYSQAAASVLVGVGDIKTLMLDSQDKYF 77
Db 236 KTVDLASDATSKSLDVVDKIDSLPIPKKT-----LNDTKLSSDLKKFLEDTNLD 288
QY 78 EATQTV-----YEWCGVATQLAAYI-----LLFEYNKKKASQAQXDLIKV 119
Db 289 ELSPLIKSLDLNLMVDLSSASSLTNLIDAVNGSGSEDPVKPLIDLNLSEKLSNLQS--LNDT 346
QY 120 LDGITHKLNBAQKSLVSSQSFNNAKGLALASQLTN--DFEKSYSYFOSVDKIRKEA 177
Db 347 LVDFLTKNLQ-----TSNN-----RLDDVIDNLEBSSNKNIDSSISTLNDIKNKV 391
QY 178 YAGAAAGVAGPFGLLISYSIAGVGBGKLIPELKNKL-KSVQFFTTLSNTVKQAKDI 236
Db 392 ISGQQPSISA-LNNVLSLNGTGRINLINNFDSKISRPINNIF--ANSIKVANDII 446
QY 237 ---DAAKLK-----LTTISIAIGEIKTETETTRFYVD-----YDDLMLSLLK-----EA 277
Db 447 TTTDKAFKPKVEETITTSIKTSGNAOESITLIRPLPLAKGLMDLLDITLSKINSGED 506

```

|    |     |        |     |
|----|-----|--------|-----|
| QY | 278 | AKKWIN | 283 |
|    |     | :::    |     |
| Db | 507 | MKKLVS | 512 |

RESULT 48

US-10-437-963-191043

; Sequence 191043, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with Rice

;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

;; FILE REFERENCE: 38-21(53221)B

;; CURRENT APPLICATION NUMBER: US/10/437,963

;; CURRENT FILING DATE: 2003-05-14

;; NUMBER OF SEQ ID NOS: 204966

;; SEQ ID NO 191043

;; LENGTH: 815

;; TYPE: PRT

;; ORGANISM: Oryza sativa

;; FEATURE:

;; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_8739C.1.pap

US-10-437-963-191043

Query Match 7.08; Score 105.5; DB 16; Length 815;

Best Local Similarity 20.18; Pred. No. 8.5;

Matches 75; Conservative 68; Mismatches 131; Indels 99; Gaps 14;

Qy 8 KTVVKNQAJETA-----DGDALDLYNK-YLDQVTPWQTFDETIKELSRFKQBYQSAASVLV 62

Db 216 KALESVSSWESKLAENEAISLMREYKRLDQQAQK--KQMKFQSEASLLDQLS 273

Qy 63 GDIKTLMSDQKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAQKILIKVLDD 122

Db 274 STKKTVTSLEE--FREKTLAEELREEIRRL-----ESSLAQAQDDKDVLEAKLKE 323

Qy 123 GITKLEAQAQKSLVSSQSFNNASGKLLALDSQLT----- 156

Db 324 KLGDNVILQEKVLSLQSDIDNGIRIELSLSLSSKEADYRNLCSFSDQTKESLELAQAK 383

Qy 157 -----NDPSEKSYFQSQVDKIRKEAYAGAA-----GVVAGPGLIISY 196

Db 384 IQLEEEVHTRNDLSKI-----SSIDLNEELQALNSAKNEAEKSELTKDYTDLUKAS 439

Qy 197 SIAAGVVEGKLIPELKNKLSVQNFFT-TLSNTVKQ-----ANKOIDAAKILITTEIAA 249

Db 440 SEARESNSLELLEKONWIKQDGLSDALSDDSKDRENTAALNKELDATKAMLENEVAA 499

Qy 250 IGEIK-----TETTRFYVDYD-----DLMSLLKEAAKQKMNTCNEYQK 290

Db 500 VKSLRESLQSTEEALTSRSEVSKLSVELDEANRNNQDLVLIQI-----SKLQDEFNEMQE 554

Qy 291 RHGKKTLPFVPEV 303

Db 555 GLTNK-LGEVESV 566

RESULT 49

US-10-369-493-11332

;; Sequence 11332, Application US/10369493

;; Publication No. US20030233675A1

;; GENERAL INFORMATION:

;; APPLICANT: Cao, Yongwei

;; APPLICANT: Hinkle, Gregory J.

;; APPLICANT: Slater, Steven C.

;; APPLICANT: Goldman, Barry S.

;; APPLICANT: Chen, Xianfeng

;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

;; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

;; FILE REFERENCE: 38-10(52052)B

;; CURRENT APPLICATION NUMBER: US/10/369,493

;; CURRENT FILING DATE: 2003-02-28

;; PRIOR APPLICATION NUMBER: US 60/360,039

;; PRIOR FILING DATE: 2002-02-21

;; NUMBER OF SEQ ID NOS: 47374

;; SEQ ID NO 11332

;; LENGTH: 873

;; TYPE: PRT

;; ORGANISM: Methanosarcina mazei

US-10-369-493-11332

Query Match 7.08; Score 105.5; DB 14; Length 873;

Best Local Similarity 20.87; Pred. No. 9.3;

Matches 77; Conservative 41; Mismatches 117; Indels 135; Gaps 15;

Qy 5 VADKTVVKNQAJETAETADGALDLYNKYLDQV-----IPWQTFDETI-----KEL----- 47

Db 112 LTDKSVLGLANKVNDERFAIDCYRRFISMFGDVVLGIDFDKFEESLIEDKKELKVESDITD 171

Qy 48 -----SREKQBYQSAASVLVGDIKTLMSDQKYFEATQTVYEWCGVATQLLAAAY 97

Db 172 LDKALKDLAERFK-----GVIKL-----EKGFEPQ-----DPKVQLQWAI 208

Qy 98 ILLFDEYNEKKAQAQKILIKVLDDGKITKLEAQAQK-----LLVSSQSF-----NNASGKLLA 150

Db 209 DAVFDSNNNPRAITYR-----KLEIDDSNGTAVNVQTMVYGNRGNTSGTGVA 256

Qy 151 LDSQLTNDPSEKSYFQSQVDKIRKEAYAGAAAGVAGPGLIISYISIAAGVVECKLIPE 210

Db 257 F---TRNPSTGERKPFGEYLINAQGE-----DVAAGIRTPDFIDT 293

Qy 211 LKNKLSVQNFFTTSLNSTVKQANKDID-----AALKLTTEI 247

Db 294 LGNKITPEAYNQLVDICRLEBAHFKMDQDIEFTQEGKLYMLQTRTGKRTAAAVKLAIDM 353

Qy 248 AAIQEIKTETETTRFYVDYDML-----SLKEAAK-----KMINTCN--E 287

Db 354 VAESGLDKETAVTRVKAHIDILLHPRIDPNKLEVVAKGLPASPGAAGVGVFTAEVAE 413

Qy 288 YQKRHGKKTLP 297

Db 414 EMAEKGKTI 423

RESULT 50

US-10-393-602-148

;; Sequence 148, Application US/10393602

;; Publication No. US20030170714A1

;; GENERAL INFORMATION:

;; APPLICANT: Gregory Dolganov

;; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory

;; NUMBER OF SEQUENCES: 151

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Dehlinger & Associates

;; STREET: 350 Cambridge Avenue, Suite 250

;; CITY: Palo Alto

;; STATE: CA

;; COUNTRY: USA

;; ZIP: 94306

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patent in Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/10/393,602

;; FILING DATE: 19-Mar-2003

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US/08/592,126

;; FILING DATE: 26-JAN-1996

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Sholtz, Charles K.

;; REGISTRATION NUMBER: 38,615

;; REFERENCE/DOCKET NUMBER: 4600-0111

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (415) 324-0880

;; TELEFAX: (415) 324-0960

;; INFORMATION FOR SEQ ID NO: 148:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 1312 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

;; HYPOTHETICAL: NO

;; ORIGINAL SOURCE:

```

; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
; SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-10-393-602-148

Query Match      7.0%; Score 105.5; DB 14; Length 1312;
Best Local Similarity 18.8%; Pred. No. 16;
Matches 75; Conservative 58; Mismatches 96; Indels 171; Gaps 17;

Qy 10 VEVVKNAIETADGALDLYNKYLDQVLPWQTFDETIK-----ELSRPKQEY 54
Db 502 MEVISLQNEKAD--LDRLKLDQEMEQLNHHHTTRTQMEMLTQDKADKDEQIRKIKSRH 559
Qy 55 SQAASVLVGDIKTLMDSDQKYPFATQTVYEWCGVATOLLAAYLLPDEYNEXKASAKD 114
Db 560 SDELTSLLG-----YFPNKKQLEDWLHLSKSK-----EINQTR----- 591
Qy 115 ILIKVLDDGITKLNKAEQSLVSSQSFNNASGKLLALDLSOLTNDPSEK-----SSYFQS 168
Db 592 -----DRLAKLN---KELASSEQNKNHNNELKRREQLSS--YEDKLFVCGSQDFES 640
Qy 169 QVDKIRKE-----AYAGAAAGVVAGPF----- 190
Db 641 DLDRKKEIEKSKQRAMLAGATA--VYSQFITQITDENQSCCPVCQVQFTEAELOEVI 698
Qy 191 -----GLI-ISYSIAAGVVEGKLIPELNKKLS 217
Db 699 SDLQSKLRAPDKLKSTESLKKKRRDEMGLVPMRQSIID--LKEKEIPELRNKLQN 756
Qy 218 VQNFFTLSNTVKQANKDIDAQKLKLTETIAAIGEIKTETTT-----RFYVDY 266
Db 757 V-----NRDIQRLKNDIEBEQETLLGTIMPEESAKVCILTDTVTIMERFQMEI 802
Qy 267 DDLMLSLKKAKEAK-----KMINTCNE--YQKRGHKKTL 297
Db 803 KQVERKIAQQAQKLGIDLDRTVQOVNQEKQEKQKHLDTV 842
```

Search completed: January 5, 2005, 11:08:51  
Job time : 66.2944 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:51:49 ; Search time 15.9474 Seconds  
(without alignments)  
1260.042 Million cell updates/sec

Title: US-09-993-292B-28

Perfect score: 1508

Sequence: 1 MTEIVADKTVEVKNNAIETA.....TCNEYQKRGKTLFVPEV 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 1499  | 99.4        | 309    | 1     | Sequence 3, Appli |
| 2          | 1499  | 99.4        | 309    | 5     | Sequence 3, Appli |
| 3          | 124   | 8.2         | 1036   | 4     | Sequence 7736, Ap |
| 4          | 118   | 7.8         | 961    | 4     | Sequence 66, Appl |
| 5          | 118   | 7.8         | 1211   | 3     | Sequence 4820, Ap |
| 6          | 116   | 7.7         | 885    | 4     | Sequence 1660, Ap |
| 7          | 116   | 7.7         | 1129   | 4     | Sequence 8019, Ap |
| 8          | 115.5 | 7.7         | 718    | 4     | Sequence 2753, Ap |
| 9          | 114   | 7.6         | 515    | 4     | Sequence 5317, Ap |
| 10         | 113.5 | 7.5         | 1196   | 4     | Sequence 3944, Ap |
| 11         | 112   | 7.4         | 606    | 4     | Sequence 2, Appli |
| 12         | 112   | 7.4         | 631    | 4     | Sequence 11, Appl |
| 13         | 111.5 | 7.4         | 808    | 4     | Sequence 4995, Ap |
| 14         | 111   | 7.4         | 477    | 1     | Sequence 3, Appli |
| 15         | 111   | 7.4         | 477    | 1     | Sequence 3, Appli |
| 16         | 111   | 7.4         | 477    | 3     | Sequence 3, Appli |
| 17         | 110.5 | 7.3         | 924    | 4     | Sequence 18798, A |
| 18         | 109.5 | 7.3         | 1454   | 4     | Sequence 5793, Ap |
| 19         | 108.5 | 7.2         | 565    | 4     | Sequence 807, App |
| 20         | 107.5 | 7.1         | 829    | 4     | Sequence 20145, A |
| 21         | 106.5 | 7.1         | 1231   | 4     | Sequence 5150, Ap |
| 22         | 105.5 | 7.0         | 1312   | 2     | Sequence 148, App |
| 23         | 105.5 | 7.0         | 1312   | 2     | Sequence 51, Appl |
| 24         | 105.5 | 7.0         | 1312   | 4     | Sequence 148, App |
| 25         | 104   | 6.9         | 794    | 4     | Sequence 1050, Ap |
| 26         | 104   | 6.9         | 962    | 3     | Sequence 4497, Ap |
| 27         | 103.5 | 6.9         | 349    | 4     | Sequence 13939, A |

Query Match 99.4%; Score 1499; DB 1; Length 309;

#### ALIGNMENTS

##### RESULT 1

US-08-557-115-3 Application US/08557115

Sequence 3, Appli

Patent No. 5731151

GENERAL INFORMATION:

APPLICANT: King, Harold C.

APPLICANT: Sathish, Mundayoor

APPLICANT: Shimmick, Thomas M.

TITLE OF INVENTION: Regulator of Contact-Mediated Hemolysis

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.A.

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/557,115

FILING DATE: 26-JUN-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Greene, Jamie L.

REGISTRATION NUMBER: 32,467

REFERENCE/DOCKET NUMBER: 03063-0171US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 818-3700

TELEFAX: (404) 818-3799

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 309 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-557-115-3

Best Local Similarity 100.0%; Pred. No. 2.3e-134;  
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPQTFDETIKELSRFKQESQAASV 60  
Db 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPQTFDETIKELSRFKQESQAASV 60

Qy 61 LVGDIKTLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAQKDILIKVL 120  
Db 61 LVGDIKTLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAQKDILIKVL 120

Qy 121 DDGITKLNKQAQSLVSSQSFNNASGKLLALDLSQTLNDFSEKSYFQSQVDKIRKEAYAG 180  
Db 121 DDGITKLNKQAQSLVSSQSFNNASGKLLALDLSQTLNDFSEKSYFQSQVDKIRKEAYAG 180

Qy 181 AAAGVWAGPGLIISYSIAAGVVEGKLIPELKNKLSKVQNFPTLNTVKQANKDIDAAK 240  
Db 181 AAAGVWAGPGLIISYSIAAGVVEGKLIPELKNKLSKVQNFPTLNTVKQANKDIDAAK 240

Qy 241 LKLTTEIAAIGEIKTETETTRFYVDYDDMLSLKKAQKMINTCNEYQKRHGKKTILFEV 300  
Db 241 LKLTTEIAAIGEIKTETETTRFYVDYDDMLSLKKAQKMINTCNEYQKRHGKKTILFEV 300

Qy 301 P 301  
Db 301 P 301

RESULT 2  
PCT-US94-05869-3  
; Sequence 3, Application PC/TUS9405869  
; GENERAL INFORMATION:  
; APPLICANT: King, C. H.  
; APPLICANT: Sathish, Mundayoor  
; APPLICANT: Shinnick, Thomas M.  
; TITLE OF INVENTION: REGULATOR OF CONTACT-MEDIATED HEMOLYSIS  
; TITLE OF INVENTION: OF MYCOBACTERIUM TUBERCULOSIS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Suite 1200, The Candler Building  
; STREET: 127 Peachtree Street, N.E.  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303-1811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/05869  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perryman, David G.  
; REGISTRATION NUMBER: 33,438  
; REFERENCE/DOCKET NUMBER: 1414.611  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 688-0770  
; TELEFAX: (404) 688-9880  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 309 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US94-05869-3

Query Match 99.4%; Score 1499; DB 5; Length 309;  
Best Local Similarity 100.0%; Pred. No. 2.3e-134;

Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPQTFDETIKELSRFKQESQAASV 60  
Db 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPQTFDETIKELSRFKQESQAASV 60

Qy 61 LVGDIKTLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAQKDILIKVL 120  
Db 61 LVGDIKTLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAQKDILIKVL 120

Qy 121 DDGITKLNKQAQSLVSSQSFNNASGKLLALDLSQTLNDFSEKSYFQSQVDKIRKEAYAG 180  
Db 121 DDGITKLNKQAQSLVSSQSFNNASGKLLALDLSQTLNDFSEKSYFQSQVDKIRKEAYAG 180

Qy 181 AAAGVWAGPGLIISYSIAAGVVEGKLIPELKNKLSKVQNFPTLNTVKQANKDIDAAK 240  
Db 181 AAAGVWAGPGLIISYSIAAGVVEGKLIPELKNKLSKVQNFPTLNTVKQANKDIDAAK 240

Qy 241 LKLTTEIAAIGEIKTETETTRFYVDYDDMLSLKKAQKMINTCNEYQKRHGKKTILFEV 300  
Db 241 LKLTTEIAAIGEIKTETETTRFYVDYDDMLSLKKAQKMINTCNEYQKRHGKKTILFEV 300

Qy 301 P 301  
Db 301 P 301

RESULT 3  
US-09-543-681A-7736  
; Sequence 7736, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 7736  
; LENGTH: 1036  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-7736

Query Match 8.2%; Score 124; DB 4; Length 1036;  
Best Local Similarity 21.5%; Pred. No. 0.0088;  
Matches 82; Conservative 56; Mismatches 127; Indels 116; Gaps 16;

Qy 4 IVADKTVEVVKNAIETADGALDLYNKYLDQVTPQTFDETIKELSRFKQESQAASV-- 61  
Db 178 IKKQKNGENVSS--ELAKASIDILINQLVD-----TASSLNNNISAFSQQNLKGLSVLSN 230

Qy 62 -----VGDITKTLMDSQDKYFEATQTVYEWCGVATQLLAAAYILL--FDEYNEKKAQAQKD 114  
Db 231 TKHLNGVGN--KLQNLPLNDKLTGLDVT---SGILSAISAFSILSNADADGTGKAAGVE 286

Qy 115 ILIKVLDD-----GITKLNKQAQSLVSSQSFNNASGKLLALDLSQ--T 156  
Db 287 LTKVLGNVGKANSQYILAQVAGLS--TSNAGSLIASAVTLAISPLSLAIADQFKRA 345

Qy 157 NDFSEKSSYFQ-----SQVDKIRKEA-----YAGAAGVVGAGPGL 192  
Db 346 NKIEYSQRPKKFGYEGDSLLAAFRKETGDAIDASLTINTALGTISAGISAASTASLIGA 405

Qy 193 IISYSIAA--GVVEGKLIPELKNKLSKVQNFPTLNTV----- 229  
Db 406 PISALVGAITGISGILEASKSMFEHVN---RWNTIAEWKTHGKNFENGYDARHS 462

Qy 230 -----KQANKDIDAAKILTT-----ETAAIGEIKTETETTRFYVDYDDLM 270



```
Db 463 AFLEDNFKLLSQYNKEYSVERSULITQOHWDLIGELASVTNGAKTLGSKSYIDY----- 518
Qy 271 LSLKBAAKMINTCNBYQKR 291
Db 519 -----YEEGKRLKLEKPNFQK 535

RESULT 4
US-09-914-259-66
; Sequence 66, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
; US-09-914-259-66

Query Match 7.8%; Score 118; DB 4; Length 961;
Best Local Similarity 22.9%; Pred. No. 0.029;
Matches 72; Conservative 54; Mismatches 138; Indels 50; Gaps 12;

Qy 7 DKTVEVVAIETADGALDLY-NKYLQVTPMQTFDETIELSRFKQYEQAAVLVGD1 65
Db 635 DKKEEVKKTLEQHDSTVTHYKNMIREQDLQLEELKQOISTLKQNEQLQTAVTQQVSQI 694
Qy 66 K-----TLLMSODK--YFEATQTVYEWCGVATQALLAAYILLFDEYNKAS-- 110
Db 695 QQHKDQVNLKVKQKDSQHGQPTDGAQ-----MNGVQPEISR---LRBEIBELKNRE 747
Qy 111 -----AQKQITIKVLDDGITK--LNEAQSLLVSSQSFNNASGKLLALDLSQLTNDPSEK 162
Db 748 LLQSLAEKSLIENKSSQSPGTCNQSSTAGDSQIAELKQELATLKSQ--NSQSV 806
Qy 163 SSYPQSQVDKIRKBAAGAAAGVAGVPFGLIISYIAAGVVEGKL-----IPELKNK 216
Db 807 ITKLQTEKQELLQKTEAFKSAVPVGESETVIATKTTD--VEGRLSALLQETKELKNEIK 864
Qy 217 SVQNFPTTSLNTVQKANKDIDAAKLTTEIAAIGEIKTETETTRFFVYVDYDMLSLIKE 276
Db 865 ALSBERTAIKEQDSSNS-----TIAIQNEKNKLEVDITDSKKQBQDLLV--LLAD 914
Qy 277 AAKQMINTCNBYQK 290
Db 915 QDQKIFSLKNKLKE 928

RESULT 5
US-09-134-001C-4820
; Sequence 4820, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4820

Db 463 AFLEDNFKLLSQYNKEYSVERSULITQOHWDLIGELASVTNGAKTLGSKSYIDY----- 518
Qy 271 LSLKBAAKMINTCNBYQKR 291
Db 519 -----YEEGKRLKLEKPNFQK 535

Query Match 7.8%; Score 118; DB 3; Length 1211;
Best Local Similarity 19.3%; Pred. No. 0.041;
Matches 69; Conservative 53; Mismatches 112; Indels 124; Gaps 13;

Qy 5 VADKTVEVVK-----NAISTADGALDLYNKYLDQVTPMQTFDETIELSRFKQYEQAA 58
Db 183 IIEESAGVLKYYKKAESIQKLDHTEDNLNRVEDILYDLGRVEPLKEAAIAKEYKQLS 242
Qy 59 S-----VLVGDITKLLMSQD-----KYFEATQTVYEWCGVATQALLAAYILLFDEYN 105
Db 243 KEMEQSDVITVTSVDIDHYTETDNRQLDERLNLHLSQQAQKE--GQQAQINQ-----LLOKYK 296
Qy 106 EKASAKQKDLIKVLDDGITKLEAQSLLVSSQSFNNASGKLLALDLSQLTND-----F 159
Db 297 GKRQONDYD-----IEKLN---YELVKATENYEQLSGKLVLEERKKNQSETNARY 344
Qy 160 SEKSSYFQSQVDKIRKEAYAGAAAGVAGVPFGLIISYIAAGVVEGKLIPELKN----- 213
Db 345 EEELDNLESQIDSIKNEK-----AQNEKLLADLNKQKQLN 380
Qy 214 -----KLKSVQNFPTTSLNTVQKANKD-----DAKLUKLT 244
Db 381 KEVQELLESLLYISDEQHQDEKLEIKNSYITLMESEQSVVNDIRFLEHTINENEAKSRLD 440
Qy 245 TETAAGIEIKTETETTRFFVYVDYDMLSLKLEAQAQKMINTCNEYQKRGHKKTLPEVPE 302
Db 441 SRLVE-----APNQLKDIOQNTQIQKEYQS---SKSKSEKVEQ 476

RESULT 6
US-09-710-279-1660
; Sequence 1660, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; US-09-710-279-1660

Query Match 7.7%; Score 116; DB 4; Length 885;
Best Local Similarity 20.0%; Pred. No. 0.04;
Matches 63; Conservative 42; Mismatches 92; Indels 118; Gaps 12;

Qy 42 ETIKELSRFKQYEQAAAS-----VLVGDITKLLMSQD-----KYFEATQTVYEWCG 88
Db 2 EPLKEAAIAKEYKQLSKEMEQQSDVITVTSVDIDHYTETDNRQLDERLNLHLSQQAQKE--G 59
Qy 89 VATQALLAAYILLFDEYNKASAKQKDLIKVLDDGITKLEAQSLLVSSQSFNNASGKL 148
Db 60 QQAQINQ-----LLOKYKKGKQONDYD-----IEKLN---YELVKATENYEQLSGKL 103
Qy 149 LALDLSQLTND-----FSEKSSYFQSQVDKIRKEAYAGAAAGVAGVPFGLIISYIAAGV 202
Db 104 NVLEERKKNQSETNARYEEELDNLESQIDSIKNEK-----A 139
```

```
Qy 203 VEGKLIPELKN-----KLSVONFFTLTSLNTVVKQANKDI- 236
Db 140 QNEKLLAELKNQKQLNKEVQESLLYSIDQEHDEKLEIKNGYYTILMSQSDVNDIR 199
Qy 237 -----DRAKLITTEIAAIGETETETETTFYVDYDDLMLSLKEAKKMINTCNE 287
Db 200 FLEHTINEAKRGLRSLRVE-----AFNQLKDIOQNTOTQKE 239
Qy 288 YQKRGHKKTLPEVPE 302
Db 240 YQS--SKSMEKVEQ 252

RESULT 7
US-09-543-681A-8019
; Sequence 8019, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8019
; LENGTH: 1129
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8019

Query Match 7.7%; Score 116; DB 4; Length 1129;
Best Local Similarity 23.3%; Pred. No. 0.057;
Matches 67; Conservative 55; Mismatches 120; Indels 46; Gaps 13;

Qy 3 EIVADKTVVKNNAIETADGALDLYNKYLDQVTPWQTFDETIKLSRPFQYSQAASVLV 62
Db 434 KIIADWLEINQTSINKAKGSLDRAEIGIDRA-----RNASPOAERLGLBIKEADS--W 485
Qy 63 GDITKLMDSODKYFEATQTVYVCGVATQI--LAAYILLPDEYNEKASAKDILLIKVL 120
Db 486 SGKSLMLDDINKY---QKINEKSAVEVLSNLSNLAKNKTQTRQATLQK--LQRI 540
Qy 121 DGITKLMNEAKS--LLVSSQSFNNASGKLALDLSQLTNDFSEKSSYFQSQVDKIRKAY 178
Db 541 NDAENKLPALQSDVILDNPKRNEHRLTFTANSLTRHDIRQSALENMTPLERELD 600
Qy 179 AGA---AAGVAGPPGLIISYIAAGVVEGK-----LIPE--LKNKLSVONFFTL 225
Db 601 AAADDIVNKIIGAPSGIVSESLIPDGLVKRAGFTKDRTLNIPDERIKDYLESVDVNY--VM 658
Qy 226 SNTVKQANKDIDAAKLKLTTEIAAI--GEIKTETETTFYVDYDDLML 270
Db 659 ENYIRQV-----APEIETAKFGRVMDQNDQIKAIT-----EYNQLI 695

RESULT 8
US-09-540-236-2753
; Sequence 2753, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2753
; LENGTH: 718
; TYPE: PRT
```

```
; ORGANISM: M.cattarrhalis
US-09-540-236-2753

Query Match 7.7%; Score 115.5; DB 4; Length 718;
Best Local Similarity 21.0%; Pred. No. 0.033;
Matches 70; Conservative 57; Mismatches 130; Indels 77; Gaps 12;

Qy 6 ADKTVVKNNAIETADGALDLYNKYLDQVTPWQTFDETIKLSRPFQY-----YSQA 57
Db 103 ATEKLEIVAKEATQ-----DKVEKTSLSVEDIKDKAQSLQEDAAADTVALKQA 149
Qy 58 ASVLVGDITKLMDSODKYFEATQTVYVWC-----GVATOLLAAYILLPEYN--- 105
Db 150 AS---DKVETTKABAQSLKDDATQT-FESAKAVEGKVEAIKEQVLDQVDSLKDDTQDN 205
Qy 106 ---EKASAKODLILVLDGITKLMNEAKSLVSSQSF--NNASGKL-----LALD--SQ 154
Db 206 TDQOEKQTLKDXAVQAATAAKRKVEDVDVVKHTTESFKNTASGKIDEIKQAADVKTTEE 265
Qy 155 LTWDFSKSYFQSQVDKIRKAYAGAAAGVAGPPGLIISYIAAGVVEGKLIPELKNK 214
Db 266 VKSLSQKADALKSGSELKQTAQT-AANDAITEAQAAVTVSGSVAAADSAQSTAQSARDK 324
Qy 215 LKSVONFFTLTSLNTVVKQANKDIDAAKLKLTTEIAAIGETETETTFYVDYDDLMLSL 274
Db 325 L-----NOLFQKGSALDEKVOELGKFGATEKINAVSENVDLATQVI 367
Qy 275 KEAAKMMINTC-----NEYQKRHGKKTIL 297
Db 368 KEAQALQTNVAQESLQAAGAAGEYDATHEDKGL 401

RESULT 9
US-09-107-532A-5317
; Sequence 5317, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5317:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...515
; SEQUENCE DESCRIPTION: SEQ ID NO: 5317:
US-09-107-532A-5317

Query Match          7.6%; Score 114; DB 4; Length 515;
Best Local Similarity 23.3%; Pred. No. 0.027;
Matches 90; Conservative 58; Mismatches 111; Indels 128; Gaps 21;

Qy 3 EIVADKTVVVKNAIETADG-----ALDLYNKYLDQVWPQTFDETIELKLS- 48
Db 24 ELTADSPKVEKENKQPTDGTISGYALKFQPSKDL-GGFVEVITP-----EALKVDL 77
Qy 49 -----RPFQVYSQ-AAVLVGDIDITLLMDSQDKYFEATOT-----VYEWK--GVATQL 93
Db 78 SNVFLLYGHDSYKPLASVYKAGTLK-LNVDDTGLHFEAELDTTYSNDVYENISKGVVDSM 136
Qy 94 LAAYILLFDEYNEKKAQAKDILIKVLDGDTIKLNEA-----QKSLIVSSQSF--- 141
Db 137 SFGFVLGDSFDKK-----EDGTIERSIDIKALNEISVVTVPAYDSSNVQVKNKSYESF 191
Qy 142 --NNASGKLALDSQLTNDFSEKSSYFQSQVDKIRK-----BAYAGAAAGVV 186
Db 192 MSNN-----QAKQTNLSLESTSKAQKESNNMEKTLIDNEKTELGRYBYEYIRSGEVR 243
Qy 187 AGPGLLIISYIAGVVEGKLIPELKNKLKSVQNF-----FTTILSN-----TVKQAN 233
Db 244 DG-----VTTVNAAAVVPVEEIVGFDLKRGSYNLAQYATVTKVTSNGOGKYPVATNQO-- 296
Qy 234 KDIDAAKLKLTETAAIGETETETTRFYVDY-----DOLMLSLKEA 277
Db 297 -----AVLATKAEALAEIGDIDAEFTS---VDYKVEYTRAGKIALSNEVEDSAVNIQEV 348
Qy 278 ----AKKMINTCNEY-----QKRHGKKT 296
Db 349 KDQLAKLVENTDNKHMILLKLTFTKKT 375

RESULT 10
US-09-107-532A-3944
; Sequence 3944, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3944:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1196
; SEQUENCE DESCRIPTION: SEQ ID NO: 3944:
US-09-107-532A-3944

Query Match          7.5%; Score 113.5; DB 4; Length 1196;
Best Local Similarity 20.0%; Pred. No. 0.11;
Matches 67; Conservative 50; Mismatches 111; Indels 107; Gaps 12;

Qy 5 VADKTVVVK-----NAIETADG---ALDLYNKYLDQVWPQTFDETIELKLSRFX 51
Db 163 IFBETAGVLKYORKKKAQKLFETEDNLSRVQDIIHELEEQLTPLAAQSEAAKEFLRLK 222
Qy 52 QEYSQA--ASVLVGDIDITLLMDSQDKYFEATQTVVCGVATQLLAAYILLFDEYNEKKAS 110
Db 223 ETUTQDVSLLMVAEIKAKKD-----WDNQKQAL----- 251
Qy 111 AQKDILIKVLDGDTIKLNEAQSLLVSSQSNFNASGKLALDSQLTNDFSEKSSYFQSQV 170
Db 252 AKFNLEGLKLSESI-----QESESILAKQRKNAQADRLIEKNQVLLDLSEKLGKQTEGQK 307
Qy 171 DKTRKEAYAGAAAGVAGVAGPGLIISYIAGVVEGKLIPELKNKLKSVQNFFTTILSNVTK 230
Db 308 DVLQERT-----KHTQKSSQEQYTSLABAQK 333
Qy 231 QA-----NKDIDAAKLKLTETI--AAIGIKTETE-----TTRFYVDYDLDLMLSLK 275
Db 334 KVHFEKLQESLMKAAAEKETETQKAEANLIKQOELEKYQKSTKELLAEIRDQYVDMQ 393
Qy 276 EAAKGMINTCNEYQKGRH-----GRKTLFEVPEV 303
Db 394 EQA--AVGNELKYLERQYIQETAKSKQTLAKQSEV 426

RESULT 11
US-08-477-831C-2
; Sequence 2, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
; APPLICANT: TURLEY, EVA A.
; ADDRESSEE: SHOWN, ZHANG
; APPLICANT: ENTWISTLE, JOYCELYN
; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10020-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Rel. #1.0, ASCII
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/477,831C  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: PIERRI, MARGARET A.  
REGISTRATION NUMBER: 30,709  
REFERENCE/DOCKET NUMBER: SIM-10  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 606 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: /desc = "RHAMM I protein"  
US-08-477-831C-2

Query Match 7.4%; Score 112; DB 4; Length 606;  
Best Local Similarity 20.5%; Pred. No. 0.054;  
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQYSQAASV 60  
DB 228 LDNLLREKEVELEKHIAHAQAIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 284  
QY 61 LVGDIKTLMSQDKYFATQTVYWCWGATOLLAAYILLDFEYNEKASAKDILIKVL 120  
DB 285 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 334  
QY 121 DDGITKLENAQKSL--VSSQSFNNASGKLAL-DSOLTN-DFSEKSSYFQSQVDKIRKEA 177  
DB 335 ESQVKYNDTAQSLRDVTAQLESYKSTLKEIEDLKLNLTLQEKVMAEKSVEDVQOOI 394  
QY 178 YAGAAAGVAVGPGFLIISYSIAAGVVEGKLIPELKNK-----LKSQVNF-----TTLS 226  
DB 395 LTAESTNQ-----EYA-----RMVDLQNRSTLKEEIKETITSSFLEKITDLK 437  
QY 227 NTVKQANKDI-----DAAKLKLTETAAIGETETETETFRFYVDYDDL 269  
DB 438 NQLRQODEDFRKLEKGRKTAENVMTELT-----MEINKWRLLYDEL 482

RESULT 12  
US-08-477-831C-11  
Sequence 11, Application US/08477831C  
Patent No. 6429291  
GENERAL INFORMATION:  
APPLICANT: TURLEY, EVA A.  
APPLICANT: SHUEN, ZHANG  
APPLICANT: ENTWISTLE, JOYCELYN  
TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10020-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Rel. #1.0, ASCII  
CURRENT APPLICATION DATA: US/08/477,831C  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: PIERRI, MARGARET A.

REGISTRATION NUMBER: 30,709  
REFERENCE/DOCKET NUMBER: SIM-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 631 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: /desc = "RHAMM I-2a"  
US-08-477-831C-11

Query Match 7.4%; Score 112; DB 4; Length 631;  
Best Local Similarity 20.5%; Pred. No. 0.058;  
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQYSQAASV 60  
DB 253 LDNLLREKEVELEKHIAHAQAIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 309  
QY 61 LVGDIKTLMSQDKYFATQTVYWCWGATOLLAAYILLDFEYNEKASAKDILIKVL 120  
DB 310 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 359  
QY 121 DDGITKLENAQKSL--VSSQSFNNASGKLAL-DSOLTN-DFSEKSSYFQSQVDKIRKEA 177  
DB 360 ESQVKYNDTAQSLRDVTAQLESYKSTLKEIEDLKLNLTLQEKVMAEKSVEDVQOOI 419  
QY 178 YAGAAAGVAVGPGFLIISYSIAAGVVEGKLIPELKNK-----LKSQVNF-----TTLS 226  
DB 420 LTAESTNQ-----EYA-----RMVDLQNRSTLKEEIKETITSSFLEKITDLK 462  
QY 227 NTVKQANKDI-----DAAKLKLTETAAIGETETETETFRFYVDYDDL 269  
DB 463 NQLRQODEDFRKLEKGRKTAENVMTELT-----MEINKWRLLYDEL 507

RESULT 13  
US-09-543-681A-4995  
Sequence 4995, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 4995  
LENGTH: 808  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-4995

Query Match 7.4%; Score 111.5; DB 4; Length 808;  
Best Local Similarity 22.0%; Pred. No. 0.093;  
Matches 66; Conservative 55; Mismatches 108; Indels 71; Gaps 14;

QY 4 IVADKTVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQY--YQAASV 62  
DB 257 IIESEAKKIVKKEKIQTSITQIELNNAI--LASV---RTGRELHKYKQKCAFCGNILP 311  
QY 63 GDIKTLMSQDKYFATQTVYWCWGATOLLAAYILLDFEYNEKASAKDILIKVLDD 122  
DB 312 ED---LQKIDKHFNKES-----EELHE-----SLNILLSTIEK 342  
QY 123 GITKLENAQKSLLVSSQSFNNASGKLALD---SOLTNDFSEKSSYFQSQVDKIRKEAYA 179

Db 343 ELLKIPNLK--IKTIDFYFNPDNLTLNDLSDKSIDISKLSLSLKEQIEKRKSDIFT 400  
Qy 180 GAAAGVAGPGLIISYIAAGVVEGKLIPELKNKLSVQNFVFTTSLNT--VKQANKDID 237  
Db 401 -----PLEFHSSVSUVEDSINELRYSYE---KIRSKNDPTKSLNTEQVKARN--- 444  
Qy 238 AAKLKLTTAAIGETETTRFYVDYDDLMLSLKE---AAKQNTICNEYQKRGK 294  
Db 445 --ELRLHEVYKFTIDIK-----YVDERNTINKLKEKEDLEKTSINIKQNVDKRRK 494

## RESULT 14

US-08-402-217A-3  
; Sequence 3, Application US/08402217A  
; Patent No. 5587301  
; GENERAL INFORMATION:  
; APPLICANT: Hawkins, Phillip R.  
; APPLICANT: Wilde, Craig G.  
; APPLICANT: Seilhamer, Jeffrey J.  
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN  
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/402,217A  
; FILING DATE: 10-MAR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Luther, Barbara J.  
; REGISTRATION NUMBER: 33954  
; REFERENCE/DOCKET NUMBER: PP-0028US  
; TELEPHONE: 415-852-0195  
; TELEFAX: 415-852-0195  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 477 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; LIBRARY: mouse  
; CLONE: GI 53979

US-08-402-217A-3

Query Match 7.4%; Score 111; DB 1; Length 477;  
Best Local Similarity 20.5%; Pred. No. 0.047;  
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

Qy 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQEYSQAASV 60  
Db 99 LDNLREKEVELEKHAHAQAIIAQEKYNDTA--QSLRDVTAQLESVQEKYNDTAQS 155  
Qy 61 LVGDIKTLMDSQDKYFEATQTVYVCGVATQLLAAAYILLDFEYNEKASQAQDKILIKVL 120  
Db 156 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQLSLRDVTAQ-L 205  
Qy 121 DDGITKLINEAKSLL-VSSQSFNNASGKLLAL-DSQLTN-DFSEKSSYFQSOVDKIRKEA 177  
Db 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLNLTLQEKVMAEKSEVDVQOQI 265

Qy 178 YAGAAAGVAGPGLIISYIAAGVVEGKLIPELKNK-----LKSQVQNF-----TTLS 226  
Db 266 LTAESTNQ-----EYA-----RMVQDLQNRSLTKBEIKEITSFLEKIDTLK 308  
Qy 227 NTVKQANKDI-----DAAKLLTTEIAAIGETETTRFYVDYDDL 269  
Db 309 NQLRQDEDFRKQLEBKGRKTAENVMTELT-----MEINKWRLLEEL 353

## RESULT 15

US-08-700-178-3  
; Sequence 3, Application US/08700178  
; Patent No. 5783669  
; Patent No. 5783669 5700912  
; GENERAL INFORMATION:  
; APPLICANT: Hawkins, Phillip R.  
; APPLICANT: Wilde, Craig G.  
; APPLICANT: Seilhamer, Jeffrey J.  
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN  
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/700,178  
; FILING DATE: August 20, 1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/402,217  
; FILING DATE: March 10, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PP-0028-1 DIV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 477 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; LIBRARY: mouse  
; CLONE: GI 53979  
; US-08-700-178-3

Query Match 7.4%; Score 111; DB 1; Length 477;  
Best Local Similarity 20.5%; Pred. No. 0.047;  
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

Qy 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQEYSQAASV 60  
Db 99 LDNLREKEVELEKHAHAQAIIAQEKYNDTA--QSLRDVTAQLESVQEKYNDTAQS 155  
Qy 61 LVGDIKTLMDSQDKYFEATQTVYVCGVATQLLAAAYILLDFEYNEKASQAQDKILIKVL 120  
Db 156 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQLSLRDVTAQ-L 205  
Qy 121 DDGITKLINEAKSLL-VSSQSFNNASGKLLAL-DSQLTN-DFSEKSSYFQSOVDKIRKEA 177

Db 206 ESQKYNDAQSLRDVSAQLESYKSTLKEIEDLKLENTLOEKVAMAEKSVEDVQQOI 265  
Qy 178 YAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNF---TTLS 226  
Db 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEEIEKITSFLEKITDCLK 308  
Qy 227 NTVKQANKDI-----DAAKLKTTEIAAIGETETETTRFYVDYDDL 269  
Db 309 NQLRQODEDFRKQLEBKGRKTAENVTMTLT-----MEINKWRLLYEEL 353

## RESULT 16

US-08-995-654-3

; Sequence 3, Application US/08995654

; Patent No. 6025138

; GENERAL INFORMATION:

; APPLICANT: Hawkins, Phillip R.

; APPLICANT: Wilde, Craig

; APPLICANT: Seilhamer, Jeffrey

; TITLE OF INVENTION: HVALURONAN RECEPTOR EXPRESSED IN HUMAN

; TITLE OF INVENTION: UNBILICAL VEIN ENDOTHELIAL CELLS

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/995,654

; FILING DATE: December 22, 1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/700,178

; FILING DATE: August 20, 1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/402,217

; FILING DATE: March 10, 1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0028-2 DIV

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 477 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-995-654-3

Query Match 7.4%; Score 111; DB 3; Length 477;

Best Local Similarity 20.5%; Pred. No. 0.047;

Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

Qy 1 WTEIVADKTVVEVKNKAIETADGALDLYNNKLDQVLPWQTFDTIKELSRFKQVYSQAASV 60

Db 99 LDNLREKEVEKQHAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155

Qy 61 LVGDIKTLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLDFYNEKASAKOILIKVL 120

Db 156 L-RDVTVAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQSLRDVTAQ-L 205

Qy 121 DGIITKLNRAQSKLL-VSSQSFNNASGKLLAI-DSQLTN-DFSEKSSYFQSOVDKIRKEA 177  
Db 206 ESQKYNDAQSLRDVSAQLESYKSTLKEIEDLKLENTLOEKVAMAEKSVEDVQQOI 265  
Qy 178 YAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNF---TTLS 226  
Db 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEEIEKITSFLEKITDCLK 308  
Qy 227 NTVKQANKDI-----DAAKLKTTEIAAIGETETETTRFYVDYDDL 269  
Db 309 NQLRQODEDFRKQLEBKGRKTAENVTMTLT-----MEINKWRLLYEEL 353

## RESULT 17

US-09-248-796A-18798

; Sequence 18798, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 18798

; LENGTH: 924

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-18798

Query Match 7.3%; Score 110.5; DB 4; Length 924;

Best Local Similarity 18.2%; Pred. No. 0.14;

Matches 52; Conservative 59; Mismatches 98; Indels 77; Gaps 8;

Qy 26 LYNKYLDQVLPWQTFDTIKELSR-----FKQVYSQAASVILVGDITKLLMDSQDKYFEATQ 81

Db 233 LHKKYVD-----SIKDLKQDFLAFKQEAAGIINVLDLDAQLSELKQKQIDLEN 281

Qy 82 TVYEWCGVATQLLAAAYILLDFYNEKASAKOILIKVLDDGITKLNRAQSKLLVSSQSF 141

Db 282 TKND---QIKELVSEHELQIEKISKOLTEKFKLVETQLLSKSHSHVQOQFKELIAESES 338

Qy 142 NNA-----SGKLLALDSQLTNDFSEKSSYFQSOVDKIRKEAYAGAAAGVVA 187

Db 339 QQVEEELLENLKTTHAKDSARILETQSDAAKESSEDYKLT-----382

Qy 188 GPFGLIISYSIAAGVVEGKLIPELKNKLSVQNFPTLSTVTVQANKDIDAANKLUTTEI 247

Db 383 -----TSEIVNDLKSQIETLKANLKL-----EEEREIQNKKLQVSEL 421

Qy 248 AAIGEIKTETETTRFYVDYDDLMLSLKKEAKKMINTCNEYQKRHG 293

Db 422 KELKELKVE-----ELSNLLKLQQLHRLHKEIELNEQLEKHLG 459

## RESULT 18

US-328-352-5793

; Sequence 5793, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: RAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 5793

```
; LENGTH: 1454
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5793

Query Match          7.3%; Score 109.5; DB 4; Length 1454;
Best Local Similarity 20.9%; Pred. No. 0.35; Indels 131; Gaps 18;
Matches 85; Conservative 54; Mismatches 137;

Qy 3 EIVADKTVVVKNAIETADGALDLYNKYLQVWPOTFDETIKE-----LSRF 50
Db 278 QVLGASVQTLASDLIADGALVVGIGVITRAILMKSS-AAIKBGMASLTASRQASVLA 335
Qy 51 KQESQASVLA-----VGDIKTLMDSDQKYFPAAT-----QTYE 85
Db 336 QAEYAEATNALNAKAHLANVRATNAETQAK-FCGATAATRYAQAOAAVTAATNAQTAAQ 394
Qy 86 -----WCGVAT-----QLLAAYILLFDEYNE-KKASACKDILI 117
Db 395 IKLNTATSIAGRLAKGAFGLIGWAGVATLGVMLAAYSFYNNKAEAKQKLAQAKVA 454
Qy 118 KVLDDGDTKL--NBAQKSLVSSQSFNNASGKLLALDLSQLTNDF---SEKSSYFQSQVDK 172
Db 455 EKADEELKLTGNDKAKAVNDLTTFAP-NAQNKALEKSSRAVGSALIDIENYARGNREVEK 513
Qy 173 IRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNKLSQVQNFPTTSLNTVKQ- 231
Db 514 ISOEARTGT-----ISYTEAI-----ERLNKIKLPTDLYENLKQQAQY 552
Qy 232 ----ANKDIDAALKL-----TTEIAAIGBIKTETE-TTRFYVDY- 266
Db 553 DDNASKASLSAEKULRVEKVLGGNEAQAIAIOHQKQADALGNATAEAKATKALQDIQ 612
Qy 267 ----DDLMLSLKLEA-----AKQWINTCNEYQRRHGKKTILFEVPEV 303
Db 613 AKQKDSVIDSYKSGWLDKGVTYAQAANLELQAKGMSAILSKEI 659

RESULT 19
US-09-538-092-807
; Sequence 807, Application US/09538092
; Patent No. 6753114
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538.092
; CURRENT FILING DATE: 2000-03-29
; PRIOR FILING DATE: 1999-04-01
; PRIOR FILING DATE: 1999-04-01
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 807
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YPR070M
US-09-538-092-807

Query Match          7.2%; Score 108.5; DB 4; Length 566;
Best Local Similarity 23.9%; Pred. No. 0.11; Indels 75; Gaps 18;
Matches 78; Conservative 43; Mismatches 131;

Qy 17 IETADGALDLYNKY-----LDQVWPOTFDETI-KELSRFKQBYSQAAVSLV- 62
Db 8 VETLDSMLIEFLKDYKPSITLENITRLCQTLGLSEFTLELSNELSRL-----STASKIIVI 63
```

```
Qy 63 -----GDIKTLMDSDQK--YFEATQTVYEWCGVATQQLAAAYILLFDEYNEKA 109
Db 64 DVDYNNKQDRIQDVKLVLASNFDFYFNORDEGEKSNILLNSLTQYDPDKAFHNNLKF 123
Qy 110 SAQKDLILKVLDDGDTKLN-EAOKSLVSSQSFNNASGKLLALDLSQLTNDFSEKSSY- 165
Db 124 LYLDDAYSHIESDSTSHNNGSSDKSLDSSNASFNN-QGKL-----DLPKYFTELSHVIRQ 177
Qy 166 -FOSQV--DKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELK-----NKLKS 217
Db 178 CFQDNCCDFKVRTN-----LNDKFGI---YILTQG-INGKEVPLAKIYLEENKSDS 224
Qy 218 VQNFFFITL--SNTVQKANKDID--AAKLLTTLTIAAIGETETETTRFYVDY----- 266
Db 225 QYRFYIYISQETKSWINESAENFSNGISLWMEIVANAKESNYTDLIWFDPEDFISPELII 284
Qy 267 DDLMLSLKLEAAKAMIN--TCNEYQKR 291
Db 285 DKVTCSSNSSSPPIIDLFNNYNSR 311

RESULT 20
US-09-248-796A-20145
; Sequence 20145, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20145
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20145

Query Match          7.1%; Score 107.5; DB 4; Length 829;
Best Local Similarity 19.5%; Pred. No. 0.23;
Matches 66; Conservative 57; Mismatches 118; Indels 97; Gaps 12;

Qy 21 DGALDL---YNKYLDQVWPOTFDETIKELSRFKQBYSQAAVSLVGDIKTLMDSDQKYF 77
Db 96 DGSINVTCTKRFELTRQVLYKSF-ERLKENLVDESMTIQAYILCSWFLF-----QQNYE 150
Qy 78 EATQTVYEWCGVATQQLAAAYILLFDEYNEKKAQAQKDLIKVLDDGDTKLNKAEQKSLVS 137
Db 151 ECWKMPFHACISISFSLGHIMNQFRETN-----TKENKPKPVVDA 191
Qy 138 SQSFNNASGKLLALDLSQLTNDFSEKSSYFOSQVDKTRKEAYAGA-----AAGVV 186
Db 192 TMGESDSS-----SEKDTMELEEDKNEDEELNAQQVRLYALKYLTSVI 237
Qy 187 AGPFGLLIISYSIAAGVVEGKLIPELKNK-----LKSQNFPTTSLNTVQKANKDIAA- 239
Db 238 CSIFGRNPISVKVGMIDSTPLNQIDQKLHVLKSESELRLNLSNLIENYMLIDISFEN 297
Qy 240 ----KLKLTTETIAAI-----GEI-----KTETETTRFYVDYDDL- 270
Db 298 VMTLKULKFDITLLEDSYDIHGKITCSLIGSEQQKNSNDNANKWIGYDDYTKGSLTT 357
Qy 271 --LSLKEAAKAMIN-----CNEYQKRHGKKTLPF 299
Db 358 KELDVLSDTITLHINSVKLLEPFVKKYETQKGKDTLSE 395

RESULT 21
```

US-09-107-532A-5150  
; Sequence 5150, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5150:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1231 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (B) LOCATION 1...1231  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5150:  
US-09-107-532A-5150  
Query Match 7.1%; Score 106.5; DB 4; Length 1231;  
Best Local Similarity 23.1%; Pred. No. 0.52;  
Matches 78; Conservative 58; Mismatches 125; Indels 77; Gaps 18;  
QY 6 ADKTVEVV---KNAIETADGALDLYNK-----YLDQVTPWQTFDE-----TIK 45  
DB 266 ADAAEEVIVSGVAQITQETIDLYNANGKRTGFLNHLHPFPVETFPQKLPHTVKTIA 325  
QY 46 ELSRFKQEYQASVINGDIKTLMDSQ-----DKYFEATQVYEWCGVATQLAAAYI 98  
DB 326 VLDRSKFPGAGGEPILL-DVQSALYDELPAVIGRYGLGSKDV-----TPQISA--- 376  
QY 99 LLFPEYNEKKAQAKDILIKVLDDGITKLNEAKSL-LVSSQSFN-----NAS 145  
DB 377 -VFDELKKDPSIRKRTFIVDDVTYQSLPEKESLDLTFPQTFQAKFWGFGSDGVTGAN 435  
QY 146 GKLLALDSQLTNDFSEKSSYFQSQ-----VDKIR-----KEAYAGAAAGVVA--GPFG 191  
DB 436 KSAIKIIGDHTDKYAQGYFYDYSKSGGLTVSHLRFGDTPIRSNLYLVEHADLVACHTP-A 494  
QY 192 LIISYSIAAGVVEGKLIPELKNKLSQNFPTTSLNTVKQ--ANKDI---DAAKLKLTTE 246  
DB 495 YLHSYDLVKLPGKGIF--LLNTLWSDBQLETHPLPLKRYLAENNIRFTYINAMRLAQE 552

QY 247 IAAIGEIKTETETTRF-----YVDYDDMLSLKKEAAK 280  
DB 553 VGLGRRINTAMETAFFKLADIIPDE-VLPLLKEALK 589  
RESULT 22  
US-08-592-126-148  
; Sequence 148, Application US/08592126  
; Patent No. 5821091  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Dolganov  
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
; TITLE OF INVENTION: Polypeptides  
; NUMBER OF SEQUENCES: 151  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,126  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 148:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1312 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54  
US-08-592-126-148  
Query Match 7.0%; Score 105.5; DB 2; Length 1312;  
Best Local Similarity 18.8%; Pred. No. 0.72;  
Matches 75; Conservative 58; Mismatches 96; Indels 171; Gaps 17;  
QY 10 VEVVNAIETADGALDLYNKYLDQVTPWQTFDETIK-----ELSRFKQEY 54  
DB 502 MEVTSIQNEKAD--LDRTLRLKQEMEQLEMLTTRTQEMLETKADKDEQIRKIKSRH 559  
QY 55 SQAASVLVGDIKTLMDSQKQPEATQVYEWCGVATQLAAAYILLFDEYNEKKAQAKD 114  
DB 560 SDELTSLUG-----YFENKQLEDLHLSKSK-----EINQTR----- 591  
QY 115 ILIKVLDDGITKLNEAKSLVSSQSFNNASGKLALDLSQLTNDFSEK-----SSYFQS 168  
DB 592 -----DRLAKLN--KELASSEQNKNHINNELKREEQUSS--YEDKLPDVGCSQDFES 640  
QY 169 QVDKIRKE-----AYAGAAAGVAGP----- 190  
DB 641 DLDRKLEIEKSKQRAMLAGATA--VYSQFIQTLDENQSCCPVCQVFQTEAELOEVI 698  
QY 191 -----GLI--ISYSIAAGVVEGKLIPELKNKLS 217  
DB 699 SDLQSKRLAPDLKSTSELKKEKRRDEMLGLVPMRQSIID--LKEKEIPELRNKLQN 756



QY 218 VQNFPTTSLNTVQKANDIDAAKLUKLTTEIAAIGETKTTETTT-----RFVVDY 266  
Db 757 V-----NRDIQRLKNDIEEQETLLGTIMPEESAKVCLTDVTIMERFQOMEL 802  
QY 267 DDLMLSLLEAKA-----KMINTCNE--YQKRHGKKTLL 297  
Db 803 KDVVERKIAQAQKLOGIDLDRTVQVQVQKQKQKHLKLDTV 842

RESULT 23  
US-08-687-080-51  
; Sequence 51, Application US/08687080  
; Patent No. 5965427  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Dolganov  
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 175  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/687,080  
; FILING DATE: 17-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/592,126  
; FILING DATE: 26-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0111.30  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1312 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NT.  
; INDIVIDUAL ISOLATE: 389 TO 4324  
US-08-687-080-51

Query Match 7.0%; Score 105.5; DB 2; Length 1312;  
Best Local Similarity 18.8%; Pred. No. 0.72;  
Matches 75; Conservative 58; Mismatches 96; Indels 171; Gaps 17;

QY 10 VEVVKNAIETADGALDLYNKYLQVLPWQTFDETIK-----ELSRFKQY 54  
Db 502 MEVISLQNEKAD--LDRTLKLDQEMEQLNHHTTTRTQMEMLTKDKADKDEQIRKIKSRH 559  
QY 55 SQASVLVAGDIKTLMDSDQKYEATQTVVEMCGVATQLLAAYILLPDEYNEKKASAKD 114  
Db 560 SDELTSLG-----YFNNKKQLEDWHLKSK-----EINQTR----- 591  
QY 115 ILIKVLDGDKTKLNEAKQSLVSSQFNASGKLLALDLSQITNDFSEK-----SSYFQS 168  
Db 592 -----DLAKUN---KELASSEQNKHNINNELKRKEEQQLSS-YEDKLPDVCQSQDFES 640

QY 169 QVDKIRKE-----AVAGRAAGVVAGPF----- 190  
Db 641 DLRLKEEIEKSKQRAMLAGATA--VYSQFITQLTDENOSCCPVQORVFQTEAELOEVI 698  
QY 191 -----GLI-ISYSIAAGVVEGKLIPKLNKLS 217  
Db 699 SDLOSGLRLAPDKLKSTESLKKKKRDEMLGLVPMRQSIID--LKEKEIPELRLNKLQN 756  
QY 218 VQNFPTTSLNTVQKANDIDAAKLUKLTTEIAAIGETKTTETTT-----RFVVDY 266  
Db 757 V-----NRDIQRLKNDIEEQETLLGTIMPEESAKVCLTDVTIMERFQOMEL 802  
QY 267 DDLMLSLLEAKA-----KMINTCNE--YQKRHGKKTLL 297  
Db 803 KDVVERKIAQAQKLOGIDLDRTVQVQVQKQKQKHLKLDTV 842

RESULT 24  
US-09-168-595-148  
; Sequence 148, Application US/09168595  
; Patent No. 6555666  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Dolganov  
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
; TITLE OF INVENTION: Polypeptides  
; NUMBER OF SEQUENCES: 151  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/168,595  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/592,126  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 148:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1312 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54  
US-09-168-595-148

Query Match 7.0%; Score 105.5; DB 4; Length 1312;  
Best Local Similarity 18.8%; Pred. No. 0.72;  
Matches 75; Conservative 58; Mismatches 96; Indels 171; Gaps 17;

QY 10 VEVVKNAIETADGALDLYNKYLQVLPWQTFDETIK-----ELSRFKQY 54  
Db 502 MEVISLQNEKAD--LDRTLKLDQEMEQLNHHTTTRTQMEMLTKDKADKDEQIRKIKSRH 559  
QY 55 SQASVLVAGDIKTLMDSDQKYEATQTVVEMCGVATQLLAAYILLPDEYNEKKASAKD 114

Db 560 SDELTSLLG-----YFPNKKQLEDWHLKSK-----EINQTR----- 591  
Qy 115 ILIKVLDGITYKLEAOKSLVSSQSFNNAAGKLLALDSQTLNDFSEK-----SSYFOS 168  
Db 592 -----DRLAKLN--KELASSEQNKHNINNELKRREQLSS-YEDKLFVDCGSDFFS 640  
Qy 169 QVDKIRKE-----AYGAAAGVVGAPF----- 190  
Db 641 DLRLKEIEKSKQORAMLAGATA--VYSQFITQLTDENQSCCPVQCFVQTAEAEQEV 698  
Qy 191 -----GLI-LSYSIAAGVVEGKLIPELKNKJLS 217  
Db 699 SDLOSKURLAPDKLSTESSELKKEKKEKDEMLGLVPMRQSLID--LKEKEIPELKNKJLN 756  
Qy 218 VQNFPTLSTVTKQANKDIDAOKLKLTEIAAIGIKTETTT-----RYVDY 266  
Db 757 V-----NRDIQRLKNDIEBQETLLGTIMPEBESAKVCLDTVTIMERFQME 802  
Qy 267 DDLMLSLLEKAAK-----KWIINTCNE--YQKRHGKKT 297  
Db 803 KDVETKIAQAQAKLQGLIDRLTVQOVNQEKQKQKHLDTV 842

RESULT 25  
US-09-710-279-1050  
; Sequence 1050, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; PRIOR FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1050  
; LENGTH: 794  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-1050

Query Match 6.9%; Score 104; DB 4; Length 794;  
Best Local Similarity 17.0%; Pred. No. 0.47; Indels 148; Gaps 14;  
Matches 69; Conservative 67; Mismatches 123; Indels 148; Gaps 14;  
Qy 9 TVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQBYSQAAVSL--VGDIK 66  
Db 189 TINKIKSAVYQANNSLPKINQFADKII-----ELNKHQDDLDAYANQFRSLGKYK 238  
Qy 67 TLMDSDQKYFEATQTVYEWCGVATOLLAAAYILLFDEY-----NEKKAQK- 113  
Db 239 GNVLDAGEKLNANVSSI-----PALNERAKLILALDSYMPNTERILNVAANDVPAQFPRI 293  
Qy 114 ----DILIKVLDGITYKLEAOKSLVSSQS-----FNNASGKLLALDSQ----- 154  
Db 294 NRGVDIASEGIDAASQGLNDA--KGYLTQAKARVGDYQEAAGRAQDVNNQANQLRQHTST 352  
Qy 155 -----LTN 157  
Db 353 TPQSAIKSHSEKSHSHSIKTVFVSQSGENQPVYGDNLNSDVKSMNTALTEALLSLN 412  
Qy 158 DFSEKSSYFQSDVKIRKEAYAGAAAGVAG-----PFGLIISYIAAGVVEGKLIP 209  
Db 413 QTDQQAQATQODIKSLKNIAIY-----GVIASDKPSEFKPELKNIKSRLENASKYNOQFID 467  
Qy 210 -----ELKNKLSQVNFPTLSTVTKQANKDIDA-----AKLKLTTETIAAIGE 252  
Db 468 ILSELEKSEHVDLSNEIKQVKENNSINDNLKSTNQLIDALSNGSSGQLEAVNVLRDLPN 527

Qy 253 IKTTETTRFYVDYDDLMLSLLEKAAKMKMINTCNEY--QKRHGKKT 297  
Db 528 LNKRLDTLRNVIK-----KELNRNLLAVSNEITDQLNKGONTL 565  
RESULT 26  
US-09-134-001C-4497  
; Sequence 4497, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4497  
; LENGTH: 962  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4497  
Query Match 6.9%; Score 104; DB 3; Length 962;  
Best Local Similarity 17.0%; Pred. No. 0.63; Indels 148; Gaps 14;  
Matches 69; Conservative 67; Mismatches 123; Indels 148; Gaps 14;  
Qy 9 TVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQBYSQAAVSL--VGDIK 66  
Db 199 TINKIKSAVYQANNSLPKINQFADKII-----ELNKHQDDLDAYANQFRSLGKYK 248  
Qy 67 TLMDSDQKYFEATQTVYEWCGVATOLLAAAYILLFDEY-----NEKKAQK- 113  
Db 249 GNVLDAGEKLNANVSSI-----PALNERAKLILALDSYMPNTERILNVAANDVPAQFPRI 303  
Qy 114 ----DILIKVLDGITYKLEAOKSLVSSQS-----FNNASGKLLALDSQ----- 154  
Db 304 NRGVDIASEGIDAASQGLNDA--KGYLTQAKARVGDYQEAAGRAQDVNNQANQLRQHTST 362  
Qy 155 -----LTN 157  
Db 363 TPQSAIKSHSEKSHSHSIKTVFVSQSGENQPVYGDNLNSDVKSMNTALTEALLSLN 422  
Qy 158 DFSEKSSYFQSDVKIRKEAYAGAAAGVAG-----PFGLIISYIAAGVVEGKLIP 209  
Db 423 QTDQQAQATQODIKSLKNIAIY-----GVIASDKPSEFKPELKNIKSRLENASKYNOQFID 477  
Qy 210 -----ELKNKLSQVNFPTLSTVTKQANKDIDA-----AKLKLTTETIAAIGE 252  
Db 478 ILSELEKSEHVDLSNEIKQVKENNSINDNLKSTNQLIDALSNGSSGQLEAVNVLRDLPN 537  
Qy 253 IKTTETTRFYVDYDDLMLSLLEKAAKMKMINTCNEY--QKRHGKKT 297  
Db 538 LNKRLDTLRNVIK-----KELNRNLLAVSNEITDQLNKGONTL 575

RESULT 27  
US-09-489-039A-13939  
; Sequence 13939, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747

```

; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13939
; LENGTH: 349
; TYPE: PRN
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13939

```

```

Query Match      6.9%; Score 103.5; DB 4; Length 349;
Best Local Similarity 18.9%; Pred. No. 0.15;
Matches 61; Conservative 42; Mismatches 121; Indels 99; Gaps 10;

QY      9  TVEVKNAIETADGALDLYNKYLDOVTPMQTFD--ETIKELSRF--KOEYSQA----- 57
DB      74  SVDFTEDALVINGKTIITVYAEKRAQHIPMQAAGAEVIECTGFYTSAEKSOAHLQAGARK 133

QY      58  --ASVLVGDITKLIMDSQDKYFEATQVYEWCGVATQLLAAAYILLDFEYNEKKASAKDI 115
DB      134  VLISAPAGEMKTIYVNVNDTDLTPDDFTIISVASCTTNCLAP----- 174

QY      116  LKVLVDGITKYLNEAQKSLIVSSQSFNNASGKLIALDSQLTNDFSEKSYFQSQVDKIRK 175
DB      175  MAVVLQD-----AFGITVGTMTTI----- 193

QY      176  EAYAGAAAGVYVAGPFL-----IISYSIAAGVVGKLIPELKNKLKSVQNFTT 224
DB      194  HATGTQS--LVDGPRGKDLRASRAAENVIPHTTGAAKAIGLVIPSLUGSKLKGHAQRVPT 252

QY      225  LSNVTQK---ANKRIDAAKLTLTEIAAIGE-----IKTETETFRFYVDYDGLM 270
DB      253  KTGSVTELVSLEKVTAAEVNQAMRQAAGNESFGYTEIEIVSSDIIGSHFGSIYDATQ 312

QY      271  LSLLLKGAARKQKINTCNEYQKRHG 293
DB      313  LEIVEAGGVOLVKTVAWYDNEYG 335

```

RESULT 28  
US-09-091-117-2  
; Sequence 2, Application US/09091117  
; Patent No. 6171589  
; GENERAL INFORMATION:  
; APPLICANT: The University of Melbourne  
; TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and  
; TITLE OF INVENTION: Vaccines  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/091,117  
; FILING DATE: 12 JUNE 1998  
; PRIOR APPLICATION DATA: PCT/AU96/00803  
; FILING DATE: 13-DEC-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: AU PNT127  
; FILING DATE: 13-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WINNER, Ellen P.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: +1 303 499 8080  
; TELEFAX: +1 303 499 8089  
; INFORMATION FOR SEQ ID NO: 2:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1030 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-091-117-2

```

[illegible]

```

RESULT 29
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND NUCLEOTIDE SEQUENCES
; TITLE OF INVENTION: EPIDERMIDIS FOR D
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,966
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,770
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

```

|                       |        |  |                 |                    |
|-----------------------|--------|--|-----------------|--------------------|
| Query Match           | 6.9%;  | Score 103.5;   | DB 3;           | Length 3696;       |
| Best Local Similarity | 22.0%; | Pred. No. 5.2;   |                 |                    |
| Matches               | 71;    | Conservative 49;   | Mismatches 144; | Indels 59; Gaps 11 |
| Qy                    | 7      | DKTVEVKNIAETADGALDLYNKVLDQVTPWQTFDETIKELSRFQKESQA-ASVLVGD      | 65              |                    |
| Db                    | 564    | DDLSEQVKNDIIPSNYTLASYNKYKXKLERAQTVLDEETNTPNQRYSQIQIDLLH        | 623             |                    |
| Qy                    | 66     | KTLLM-----DSQDKYFQATQTYEWCGVATQLLAAYILLFDEYNKSKASQAOKDILIK     | 118             |                    |
| Db                    | 624    | QTTLLINVSASREINDKAQEMTDAVYDSTELTTE---EKDTLVQDIENHKNKSEISNNIDDE | 680             |                    |
| Qy                    | 119    | VLDGDGITKNEAKSKLLVSS-----QSFNNASGKLLADSLQTLNDFSEKSSYFO         | 167             |                    |
| Db                    | 681    | LTDQGVVERVKEAGLHLESTPHTVPKNARQVNNRA-----DQKTLIRNHAEATE         | 734             |                    |
| Qy                    | 168    | SQVDKIRK-BAYAGAAAGVVGAGPGLIISYSIAGVVGEGKLIPELKNKLSKVQNFPTTLS   | 226             |                    |
| Db                    | 735    | EONEAIRQVRAHSSDATA-----KIGAEETDTTVNEARDNGTKLIA---TDVP          | 779             |                    |

```

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4284
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-4284

Query Match
Best Local Similarity 17.5%; Pred. No. 0.61;
Matches 58; Conservative 54; Mismatches 130; Indels 90; Gaps 9;

QY 9 TVEVKNATETADGALDLYNKYLDQVTPWOTFDTIKELSRFKQ--EYSQAASVVLVDGIK 66
Db 220 TAHIIQSMADKADTEIIIPOTILEFGLPEFDEVIKAESEFKPAEKDREGRVLDRLDA 279
QY 67 TLLMDSQD-KYFEATQTVYEWCG-----VATQLLAAYILLFDEYNEKKAQAQKDI--- 115
Db 280 LVTIGEDARDDDAVAEKPGGGYRVVVAIAVSHVYVRLDSALNEEABERGTSVYPPH 339
QY 116 -----LIKVLDDGITKLINEAQKSLVSSQSFNASGKL-----LALDSQLTNDFSEKS 163
Db 340 FVLPMPLPEALSNGLSLNPVHVRDLCMVCDLKLSTRTGVTGYEFYFPVAVHSHKARLTYYQVG 399
QY 164 SYFQSQVDKIRKEAYAGAAAGVVGPFGLIISYSIAAGVVEGKLIPELNKLKSVQNFFT 223
Db 400 QYFEGATDAIPKDR-----DIHKSILNTLFQLYQ 427
QY 224 TILSNTVQKANKIDAALKLTTEIAAIGETETETTFYDYDDI-----MISLLKEAA 278
Db 428 ILKN-----LRVDRHAMEFETIETMTFDELGGIKELIPRTRNDA 467
QY 279 KQMINTC-----NEYQKRHGKKTLPVPE 302
Db 468 HKLIEECMLLANVAAAEYALEHDIPMLRYVHE 499

RESULT 32
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5178

Query Match
Best Local Similarity 19.5%; Pred. No. 0.84;
Matches 65; Conservative 63; Mismatches 130; Indels 76; Gaps 13;

QY 11 EWKNAIETADGALDLYNKYLDQ-----VIPWOTFDTIKELSRFKQ----- 52
Db 115 ELVDNQFELRESKVNQGNQFVOLLGVNAEQFRLFPQGEFKFLQSNKDKQSILRT 174
QY 53 -----EYSQAASVVLVDGIKTLMDSDQKYFEATQTVYEWCGVAT---QLLAAY----- 97
Db 175 LFNSERFDEIRHLLVENVQEKVQIENRY---TQIENLWNDIDTFNDELALYKELESSQ 231
QY 98 -----ILLFDEYNE-----KKASAQKDIILIKVLDDGITKLINEAQKSLVSSQSFNASGK 147
Db 232 TDKMIKFPQFNDYGCCKLILKSFEEAKNKITKELDD-----LNHKKYKNVVELSENTKCLKAE 287

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4284
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-4284

Query Match
Best Local Similarity 17.5%; Pred. No. 0.61;
Matches 58; Conservative 54; Mismatches 130; Indels 90; Gaps 9;

QY 9 TVEVKNATETADGALDLYNKYLDQVTPWOTFDTIKELSRFKQ--EYSQAASVVLVDGIK 66
Db 220 TAHIIQSMADKADTEIIIPOTILEFGLPEFDEVIKAESEFKPAEKDREGRVLDRLDA 279
QY 67 TLLMDSQD-KYFEATQTVYEWCG-----VATQLLAAYILLFDEYNEKKAQAQKDI--- 115
Db 280 LVTIGEDARDDDAVAEKPGGGYRVVVAIAVSHVYVRLDSALNEEABERGTSVYPPH 339
QY 116 -----LIKVLDDGITKLINEAQKSLVSSQSFNASGKL-----LALDSQLTNDFSEKS 163
Db 340 FVLPMPLPEALSNGLSLNPVHVRDLCMVCDLKLSTRTGVTGYEFYFPVAVHSHKARLTYYQVG 399
QY 164 SYFQSQVDKIRKEAYAGAAAGVVGPFGLIISYSIAAGVVEGKLIPELNKLKSVQNFFT 223
Db 400 QYFEGATDAIPKDR-----DIHKSILNTLFQLYQ 427
QY 224 TILSNTVQKANKIDAALKLTTEIAAIGETETETTFYDYDDI-----MISLLKEAA 278
Db 428 ILKN-----LRVDRHAMEFETIETMTFDELGGIKELIPRTRNDA 467
QY 279 KQMINTC-----NEYQKRHGKKTLPVPE 302
Db 468 HKLIEECMLLANVAAAEYALEHDIPMLRYVHE 499

RESULT 31
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5178

Query Match
Best Local Similarity 19.5%; Pred. No. 0.84;
Matches 65; Conservative 63; Mismatches 130; Indels 76; Gaps 13;

QY 11 EWKNAIETADGALDLYNKYLDQ-----VIPWOTFDTIKELSRFKQ----- 52
Db 115 ELVDNQFELRESKVNQGNQFVOLLGVNAEQFRLFPQGEFKFLQSNKDKQSILRT 174
QY 53 -----EYSQAASVVLVDGIKTLMDSDQKYFEATQTVYEWCGVAT---QLLAAY----- 97
Db 175 LFNSERFDEIRHLLVENVQEKVQIENRY---TQIENLWNDIDTFNDELALYKELESSQ 231
QY 98 -----ILLFDEYNE-----KKASAQKDIILIKVLDDGITKLINEAQKSLVSSQSFNASGK 147
Db 232 TDKMIKFPQFNDYGCCKLILKSFEEAKNKITKELDD-----LNHKKYKNVVELSENTKCLKAE 287

```

```

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4284
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-4284

Query Match
Best Local Similarity 17.5%; Pred. No. 0.61;
Matches 58; Conservative 54; Mismatches 130; Indels 90; Gaps 9;

QY 9 TVEVKNATETADGALDLYNKYLDQVTPWOTFDTIKELSRFKQ--EYSQAASVVLVDGIK 66
Db 220 TAHIIQSMADKADTEIIIPOTILEFGLPEFDEVIKAESEFKPAEKDREGRVLDRLDA 279
QY 67 TLLMDSQD-KYFEATQTVYEWCG-----VATQLLAAYILLFDEYNEKKAQAQKDI--- 115
Db 280 LVTIGEDARDDDAVAEKPGGGYRVVVAIAVSHVYVRLDSALNEEABERGTSVYPPH 339
QY 116 -----LIKVLDDGITKLINEAQKSLVSSQSFNASGKL-----LALDSQLTNDFSEKS 163
Db 340 FVLPMPLPEALSNGLSLNPVHVRDLCMVCDLKLSTRTGVTGYEFYFPVAVHSHKARLTYYQVG 399
QY 164 SYFQSQVDKIRKEAYAGAAAGVVGPFGLIISYSIAAGVVEGKLIPELNKLKSVQNFFT 223
Db 400 QYFEGATDAIPKDR-----DIHKSILNTLFQLYQ 427
QY 224 TILSNTVQKANKIDAALKLTTEIAAIGETETETTFYDYDDI-----MISLLKEAA 278
Db 428 ILKN-----LRVDRHAMEFETIETMTFDELGGIKELIPRTRNDA 467
QY 279 KQMINTC-----NEYQKRHGKKTLPVPE 302
Db 468 HKLIEECMLLANVAAAEYALEHDIPMLRYVHE 499

RESULT 32
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5178

Query Match
Best Local Similarity 19.5%; Pred. No. 0.84;
Matches 65; Conservative 63; Mismatches 130; Indels 76; Gaps 13;

QY 11 EWKNAIETADGALDLYNKYLDQ-----VIPWOTFDTIKELSRFKQ----- 52
Db 115 ELVDNQFELRESKVNQGNQFVOLLGVNAEQFRLFPQGEFKFLQSNKDKQSILRT 174
QY 53 -----EYSQAASVVLVDGIKTLMDSDQKYFEATQTVYEWCGVAT---QLLAAY----- 97
Db 175 LFNSERFDEIRHLLVENVQEKVQIENRY---TQIENLWNDIDTFNDELALYKELESSQ 231
QY 98 -----ILLFDEYNE-----KKASAQKDIILIKVLDDGITKLINEAQKSLVSSQSFNASGK 147
Db 232 TDKMIKFPQFNDYGCCKLILKSFEEAKNKITKELDD-----LNHKKYKNVVELSENTKCLKAE 287

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4284
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-4284

Query Match
Best Local Similarity 17.5%; Pred. No. 0.61;
Matches 58; Conservative 54; Mismatches 130; Indels 90; Gaps 9;

QY 9 TVEVKNATETADGALDLYNKYLDQVTPWOTFDTIKELSRFKQ--EYSQAASVVLVDGIK 66
Db 220 TAHIIQSMADKADTEIIIPOTILEFGLPEFDEVIKAESEFKPAEKDREGRVLDRLDA 279
QY 67 TLLMDSQD-KYFEATQTVYEWCG-----VATQLLAAYILLFDEYNEKKAQAQKDI--- 115
Db 280 LVTIGEDARDDDAVAEKPGGGYRVVVAIAVSHVYVRLDSALNEEABERGTSVYPPH 339
QY 116 -----LIKVLDDGITKLINEAQKSLVSSQSFNASGKL-----LALDSQLTNDFSEKS 163
Db 340 FVLPMPLPEALSNGLSLNPVHVRDLCMVCDLKLSTRTGVTGYEFYFPVAVHSHKARLTYYQVG 399
QY 164 SYFQSQVDKIRKEAYAGAAAGVVGPFGLIISYSIAAGVVEGKLIPELNKLKSVQNFFT 223
Db 400 QYFEGATDAIPKDR-----DIHKSILNTLFQLYQ 427
QY 224 TILSNTVQKANKIDAALKLTTEIAAIGETETETTFYDYDDI-----MISLLKEAA 278
Db 428 ILKN-----LRVDRHAMEFETIETMTFDELGGIKELIPRTRNDA 467
QY 279 KQMINTC-----NEYQKRHGKKTLPVPE 302
Db 468 HKLIEECMLLANVAAAEYALEHDIPMLRYVHE 499

RESULT 31
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5178

Query Match
Best Local Similarity 19.5%; Pred. No. 0.84;
Matches 65; Conservative 63; Mismatches 130; Indels 76; Gaps 13;

QY 11 EWKNAIETADGALDLYNKYLDQ-----VIPWOTFDTIKELSRFKQ----- 52
Db 115 ELVDNQFELRESKVNQGNQFVOLLGVNAEQFRLFPQGEFKFLQSNKDKQSILRT 174
QY 53 -----EYSQAASVVLVDGIKTLMDSDQKYFEATQTVYEWCGVAT---QLLAAY----- 97
Db 175 LFNSERFDEIRHLLVENVQEKVQIENRY---TQIENLWNDIDTFNDELALYKELESSQ 231
QY 98 -----ILLFDEYNE-----KKASAQKDIILIKVLDDGITKLINEAQKSLVSSQSFNASGK 147
Db 232 TDKMIKFPQFNDYGCCKLILKSFEEAKNKITKELDD-----LNHKKYKNVVELSENTKCLKAE 287

```

Qy 148 LLAIDSLQTNDFSEKSSYFQSQVDKIRKEAYAGAAAGVAGPFF-----GLIISYS 197  
Db 288 KIKED-----DLKKEQNY-----IDKLKQELKMIQESKVLITYFTRLOSLKDKDELVSLLH 338  
Qy 198 IAAAGVGEKLPKLNKLSQVNFPTTSLNTVQAKNDIDAALKLTTEIAAIGEIK-- 254  
Db 339 EOSKLNETHYNEIKGQKQLEH--LSTRENEITQFNQVLEKNQVFFNQDLKLISSYQOKP 397  
Qy 255 -TETTRFYVDYDMLMLSLKEAAKMWNTCNE 287  
Db 398 VIEBEIKRLYSEYNDLITK--KEELTKEMNNKNK 429

## RESULT 33

US-09-248-796A-20613  
; Sequence 20613, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 20613  
; LENGTH: 722  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-20613

Query Match 6.8%; Score 102; DB 4; Length 722;  
Best Local Similarity 20.0%; Pred. No. 0.63;  
Matches 51; Conservative 51; Mismatches 99; Indels 54; Gaps 9;

Qy 38 QTFDETIKELSRFQKQYSQAASVLVGDITKLLMDSQDKYFEATQ-----TVYEWCGVATQ 92  
Db 211 QIQYDITKTLKSTREELNGSKTEIL--RLKALLRSEBELVQVKENYKTSVHDYEQDLAQ 269  
Qy 93 LLAAYILLDFBYNEKKASAKDILIKVLDDGITKLEAQSLLVSSQSFNNASGKLALD 152  
Db 270 LKVKHETLLS--RNKDINESLEIYKGRSDEYKKLELAESAIAISKRHEQATKEMKESR 327  
Qy 153 SQ-----LTDNFESEKSSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYIA 199  
Db 328 SQQLLVREELRTTQILKDFRIKVENLEATIE-----KNHOLD 366  
Qy 200 AGVVEGKLIP--LKNKLSQVQ--FFTTLSNTVQAKNDIDAALKLTTEIAAIGEIKTE 256  
Db 367 ANKEEIKQIQKLNLYHLKNFENKELNEKLEKEIKNLNRLD-----FKTDI-----ETKLI 417  
Qy 257 TETTRFYVDYDML 271  
Db 418 KENKKQLDYEDVLL 432

## RESULT 34

US-09-134-001C-3159  
; Sequence 3159, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3159  
; LENGTH: 10182  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3159

Query Match 6.8%; Score 102; DB 3; Length 10182;  
Best Local Similarity 20.7%; Pred. No. 33;  
Matches 65; Conservative 47; Mismatches 104; Indels 98; Gaps 12;

Qy 9 TVEVVKNAIETADGALDLYNKYLDQVIPMOTFDETIKELSRFQKQYSQAASVLVGDITKL 68  
Db 4407 TVEQVKESVANA-----QQVI-----QDLQARTSLVPD-KTQ 4438  
Qy 69 LMDSDQKYFEATQTVYEWCGVATQLLAAAYILLDFEYNEKASAKDILIKVLDDGITKLN 128  
Db 4439 LQEAKNRLSENSINOQDTDTGMDQDSL-----NNYNDKLAKARQNL-----EKISKVL 4485  
Qy 129 EAQKSLIVSSQSFNNASGKLALD---SOLT-----NDFSEKSSYFQSQVDKIRKEAY 178  
Db 4486 GGQPTVABIRQNTDEANAHKQALDTARSQTLNREPYNHNNESHNLNNAQKONFK---- 4541  
Qy 179 AGAAAGVAGPFGLLIISYSIAAGVVEGKLIPKLNKLSQVNFPTTSLNTV-----KQA 232  
Db 4542 ----AQVNSAP-----NHNLTETIKRKAADTLNQSMTALSESIAADYENQKQQ 4583  
Qy 233 NKDIDAALKK---LTTEIAAIGEIKTETETTRFYVDYDD-----LMLS 273  
Db 4584 ENYLDASNKRQDYDNVNAAKGILNQTSPTMSADVIDOKAEDVKKTKTALDGNQRLV 4643  
Qy 274 LKEAAKKMWNTCNE 287  
Db 4644 AKQOALNHLNLTND 4657

## RESULT 35

US-09-492-709A-302  
; Sequence 302, Application US/09492709A  
; Patent No. 6720139  
; GENERAL INFORMATION:  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Trawick, John  
; APPLICANT: Forsyth, R. Alllyn  
; APPLICANT: Froelich, Jamie M.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
; FILE OF INVENTION: ESCHERICHIA COLI  
; FILE REFERENCE: ELITFA.001A  
; CURRENT APPLICATION NUMBER: US/09/492,709A  
; CURRENT FILING DATE: 2000-01-27  
; NUMBER OF SEQ ID NOS: 485  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 302  
; LENGTH: 2383  
; TYPE: PRT  
; ORGANISM: E. Coli  
US-09-492-709A-302

Query Match 6.7%; Score 101; DB 4; Length 2383;  
Best Local Similarity 22.9%; Pred. No. 4.7;  
Matches 61; Conservative 51; Mismatches 102; Indels 52; Gaps 11;

Qy 41 DETIKELSRFQKQYSQAASV--LVGDIKLLMDSQDKYFEATQTVY--EWCGVATQLL--- 94  
Db 685 DENDKPVKEQKQQLNNAVSIDNVKPGVTTDWTADGVYKATYATYKSGSLTAKLLMQN 744

QY 95 -----AAVILFDEYNEKA-----SAQKILIKVLDGKITKLEAKSL 134  
 Db 745 WNEHLTPAGFIIDANPQSAKIATLSASNGVLANENAAVTSVNVADEGSNFINDHVTTF 804  
 QY 135 LV-----SSQSFNNSAGKLLALDSQITNDFSEKSVFQSDVKIRKAYAGAAAGVAGPFG 191  
 Db 805 AVLSGSATSFNNQNTAKTDVNGLAT--FDLKSSKQEDNTEVTELE-----NGVKQ--T 853  
 QY 192 LIISVSTAAGVEGKLIPELKNKLKSVQNFPTLSTNTVKQK-----DIDAACL 241  
 Db 854 LIVSFVGSSTAQVDL--QKSNVAVADGNSVTMTATVRDAKGNLLNDVMVTNNSAEA 912  
 QY 242 KLT-TEIAAIGIKETETTFYDY 266  
 Db 913 KLSQTEVNSHDGIATATLTLKNGDY 938

RESULT 36  
 US-08-434-000A-4  
 ; Sequence 4, Application US/08434000A  
 ; Patent No. 6046037  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ANDREW C. HIATT, JULIAN  
 ; APPLICANT: K.-C. MA, THOMAS LEHNER  
 ; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: Suite 4700  
 ; CITY: Los Angeles  
 ; STATE: California  
 ; COUNTRY: U.S.A.  
 ; ZIP: 90071  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; MEDIUM TYPE: storage  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0  
 ; SOFTWARE: Word Perfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/434,000A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA: including application 1  
 ; PRIOR APPLICATION DATA: described below:  
 ; APPLICATION NUMBER: 08/367,395  
 ; FILING DATE: 12/30/94  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Guise, Jeffrey W.  
 ; REGISTRATION NUMBER: 34,613  
 ; REFERENCE/DOCKET NUMBER: 212/127  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 552-8400  
 ; TELEFAX: (619) 552-0159  
 ; TELEX: 67-3510  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 746 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; DESCRIPTION: Human Polymunoglobulin Receptor  
 ; US-08-434-000A-4

Query Match 6.7%; Score 100.5; DB 3; Length 746;  
 Best Local Similarity 22.6%; Pred. No. 0.92;  
 Matches 45; Conservative 41; Mismatches 70; Indels 43; Gaps 11;  
 QY 86 WCGVAT-----QLLAAYILLFDEYNEKASAKQDILI-----KVLDGKITKLEAKSL 131

Db 525 WCGVKQGHFYGETAAVTVAV-----EERKAGSRDVSIAKADAPDEKVLDSG---FREIE 577  
 QY 132 KSLIVSSQSFNNSAGKLLALDSQITNDFSEKSVFQSDVKIRKAYAGAAAGVAG--P 189  
 Db 578 NKATQDPRLF--AEKAVA-----DTRQADGSRASVDSGSSEGGSSRALVSTLP 628  
 QY 190 FGLIISY-STAAGVEGKLIPELKNKLK-SVQNFPTLSTNTVKQ-----ANKDIDAACL 241  
 Db 629 LGLVLAVGAVAGVARAR--HRKNVDRVSTIRSYRTDISMSDFENSREFGANDNMGASSI 685  
 QY 242 KLTTEIAAIGIKETETTT 260  
 Db 686 TQETSLGGKEEFVATTEST 704

RESULT 37  
 US-09-312-157-4  
 ; Sequence 4, Application US/09312157  
 ; Patent No. 6303341  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ANDREW C. HIATT, JULIAN  
 ; APPLICANT: K.-C. MA, THOMAS LEHNER  
 ; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: 633 West Fifth Street  
 ; CITY: Suite 4700  
 ; STATE: California  
 ; COUNTRY: U.S.A.  
 ; ZIP: 90071  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; MEDIUM TYPE: storage  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0  
 ; SOFTWARE: Word Perfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/312,157  
 ; FILING DATE: 14-May-1999  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/434,000  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Guise, Jeffrey W.  
 ; REGISTRATION NUMBER: 34,613  
 ; REFERENCE/DOCKET NUMBER: 212/127  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 552-8400  
 ; TELEFAX: (619) 552-0159  
 ; TELEX: 67-351  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 746 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; DESCRIPTION: Human Polymunoglobulin Receptor  
 ; US-09-312-157-4

Query Match 6.7%; Score 100.5; DB 3; Length 746;  
 Best Local Similarity 22.6%; Pred. No. 0.92;  
 Matches 45; Conservative 41; Mismatches 70; Indels 43; Gaps 11;  
 QY 86 WCGVAT-----QLLAAYILLFDEYNEKASAKQDILI-----KVLDGKITKLEAKSL 131  
 Db 525 WCGVKQGHFYGETAAVTVAV-----EERKAGSRDVSIAKADAPDEKVLDSG---FREIE 577

Qy 132 KSLVSSQSFNNASGKLLALDSQLTNDPSEKSSVFQSDVKIRKEAYAGAAAGVAG--P 189  
Db 578 NKAIQDPLFP--AEEKAVA-----DTRDQADGSRASVDSSGSEEGGSRALVSTLVP 628  
Qy 190 FGLIISY-SIAAGVVECKLIPELKNKLLK-SVONPFTTSLNTVKQ-----ANKDIDAACL 241  
Db 629 LGLVLAVGAVGVARAR---HRKNVDRVRSIRYRTDMSDPNSREFGANDMGASSI 685  
Qy 242 KLTTETIAAIGEIKTETETT 260  
Db 686 TQETSLGGKEEFVATTEST 704

## RESULT 38

US-09-315-793-52  
; Sequence 52, Application US/09315793  
; Patent No. 6221597  
; GENERAL INFORMATION:  
; APPLICANT: Roberts, Christopher J.  
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL  
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION  
; TITLE OF INVENTION: DRUGS  
; FILE REFERENCE: 9301-048  
; CURRENT APPLICATION NUMBER: US/09/315,793  
; CURRENT FILING DATE: 1999-05-21  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 52  
; LENGTH: 1093  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-315-793-52

Query Match 6.7%; Score 100.5; DB 3; Length 1093;  
Best Local Similarity 21.6%; Pred. No. 1.6;  
Matches 63; Conservative 47; Mismatches 88; Indels 93; Gaps 13;  
Qy 24 LDLYNKYLDQVWPQTFTDIKELSRPKQBYSOAASVLVGDIKT---LLMDSQDKYFEAT 80  
Db 664 INLKNEYNDRK---STLDALSNQSGYRHELSELAS-KNDDINREAHQLEIRKYYTMRK 719  
Qy 81 QTVYEWCGVATQLLAAYILLFDEYNEKKAQAQDI--LIKVLDDGI-----TKL 127  
Db 720 STIE-----TLREKLDQLKREARKDVSKIKDIDDIQOILLKQRLHLSKM 765  
Qy 128 NEAQKSL-----LVSSQ-----SFNNASGKLLALDSQLTNDPSEKSSY----- 165  
Db 766 ASSMKSLKNCKQKELISTQILQFEAQNMDVSNVDVIGFFNEREADLKSQYEDKKFVKEMR 825  
Qy 166 ----FQSQVDKIR-----KEAYAGAAAGV-VAGPFGI-----IISYSIA 199  
Db 826 DTPEFQSWREIRSYDQDTKEKLNKVAEKVEEGNFNLSFVQDVLDKLESEIAMVNHDES 885  
Qy 200 AGVVEGKLIPELKNKLSVQNFTTSLNTVKQAKDIDAAKLKTTEIAAI 250  
Db 886 AVTILQVTAELRE-----LEHTVPQOSKDLTKAKLKEDHAVL 925

## RESULT 39

US-09-538-092-701  
; Sequence 701, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR FILING DATE: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965

; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuraPatsSeqformatter Version 0.9  
; SEQ ID NO 701  
; LENGTH: 1093  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number YOL034W  
US-09-538-092-701

Query Match 6.7%; Score 100.5; DB 4; Length 1093;

Best Local Similarity 21.8%; Pred. No. 1.6;  
Matches 63; Conservative 47; Mismatches 88; Indels 93; Gaps 13;  
Qy 24 LDLYNKYLDQVWPQTFTDIKELSRPKQBYSOAASVLVGDIKT---LLMDSQDKYFEAT 80  
Db 664 INLKNEYNDRK---STLDALSNQSGYRHELSELAS-KNDDINREAHQLEIRKYYTMRK 719  
Qy 81 QTVYEWCGVATQLLAAYILLFDEYNEKKAQAQDI--LIKVLDDGI-----TKL 127  
Db 720 STIE-----TLREKLDQLKREARKDVSKIKDIDDIQOILLKQRLHLSKM 765  
Qy 128 NEAQKSL-----LVSSQ-----SFNNASGKLLALDSQLTNDPSEKSSY----- 165  
Db 766 ASSMKSLKNCKQKELISTQILQFEAQNMDVSNVDVIGFFNEREADLKSQYEDKKFVKEMR 825  
Qy 166 ----FQSQVDKIR-----KEAYAGAAAGV-VAGPFGI-----IISYSIA 199  
Db 826 DTPEFQSWREIRSYDQDTKEKLNKVAEKVEEGNFNLSFVQDVLDKLESEIAMVNHDES 885  
Qy 200 AGVVEGKLIPELKNKLSVQNFTTSLNTVKQAKDIDAAKLKTTEIAAI 250  
Db 886 AVTILQVTAELRE-----LEHTVPQOSKDLTKAKLKEDHAVL 925

## RESULT 40

US-08-328-254-6  
; Sequence 6, Application US/08328254  
; Patent No. 5710022  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Xueliang  
; APPLICANT: Lee, Wen-Hwa  
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/328,254  
; FILING DATE: 24-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/141,239  
; FILING DATE: 22-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-CJ 1191  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949

|   |   |
|---|---|
| INFORMATION FOR SEQ ID NO: 6:           |   |
| SEQUENCE CHARACTERISTICS:               |   |
| LENGTH: 2482 amino acids                |   |
| TYPE: amino acid                        |   |
| TOPOLOGY: linear                        |   |
| MOLECULE TYPE: protein                  |   |
| US-08-328-254-6                         |   |
| Query Match                             |   |
| Best Local Similarity                   |   |
| Matches                                 | 74; Conservative  |
| 6.7%; Score 100.5; DB 1; Length 2482;   | 22.0%; Pred. No. 5.6;   |
| 63; Mismatches 129; Indels 71; Gaps 16; |   |
| QY                                      | 9 TVEVYKNAIETADGALDI-----YNYKLDQVIPWQ-TFDETIKELSRFKQYSQ 56            |
| DB                                      | 1950 TLEVLQSSYKYLELELTCKDKMSFVEKVNKMTAKETELQREHMAQKTAELQBELSG 2009    |
| QY                                      | 57 AASVLVGDIKTLMD---SQKYFEAT-----QTVYEWCGVATQLLAAYIL 99               |
| DB                                      | 2010 EKNRLAGELQLLLEIKSKDKQLKELTLENSELKSLDKCHKQDQVEKEGKVRERIEAYQL 2069 |
| QY                                      | 100 LFDEYNEKKASA-----QKDILIKVLDGKITKLENAQKS-----LLVSSOSFNNAS 145      |
| DB                                      | 2070 RLHE-AEKHQALLDNTNKQYEVEIQTVREKLTSEKCLSSQKLEIDLLKSSKEELNNS- 2127  |
| QY                                      | 146 GKLLALDSQLTNDFSEKSSYFQSQVDKIRKEAYAGAAAGVAGPFGLIISYSIAAGVVEG 205   |
| DB                                      | 2128 ---LKATTQILEELKTKMDNLKYVNLKENER-----AQGKMKLLIK-SCKQLEEEK 2177    |
| QY                                      | 206 KLIPELKNKLKSVQNFFTLSNTVKQANKDIDAAKLLTTEIAAIGE-IKTETETTRFYV 264    |
| DB                                      | 2178 EILQKELSQAQAQKQKT--GTVMDTKVD-----ELTTEIKELKETLEBKTEADEYL 2229    |
| QY                                      | 265 D-YDDMLSLIK-EAAKGMINT-----CNEYQKRGHK 294                          |
| DB                                      | 2230 DKYCSLLISHEKLEKAKEMLETQVAHLCSQSQKQDSR 2266                       |

RESULT 41

US-09-538-092-1154

; Sequence 1154, Application US/09538092

; Patent No. 6753314

; GENERAL INFORMATION:

; APPLICANT: Glot, Loic

; APPLICANT: Mansfield, Traci A.

; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same

; FILE REFERENCE: 15966-542

; CURRENT APPLICATION NUMBER: US/09/538,092

; CURRENT FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/127,352

; PRIOR FILING DATE: 1999-04-01

; PRIOR APPLICATION NUMBER: 60/178,965

; PRIOR FILING DATE: 2000-02-01

; NUMBER OF SEQ ID NOS: 1387

; SOFTWARE: CuratPatSeqFormat Version 0.9

SEQ ID NO 1154

LENGTH: 3210

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc.feature

LOCATION: (0)...(0)

OTHER INFORMATION: Polypeptide Accession Number P49454

US-09-538-092-1154

Query Match

Best Local Similarity

Matches

6.7%; Score 100.5; DB 4; Length 3210;

22.0%; Pred. No. 8.2;

74; Conservative

63; Mismatches 129; Indels 71; Gaps 16;

QY

9

TVEVYKNAIETADGALDI-----YNYKLDQVIPWQ-TFDETIKELSRFKQYSQ 56

DB

2678

TLEVLQSSYKYLELELTCKDKMSFVEKVNKMTAKETELQREHMAQKTAELQBELSG 2737

QY

57

AASVLVGDIKTLMD---SQKYFEAT-----QTVYEWCGVATQLLAAYIL 99





```
Db 452 YIGPYVTGKNAFPMVN 469
;
; LENGTH: 1960
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P35579
US-09-538-092-1077

Query Match
Best Local Similarity 6.6%; Score 99.5; DB 4; Length 1960;
Matches 71; Conservative 59; Mismatches 119; Indels 95; Gaps 16;

QY 3 EIVADKTVVVKNAIETADGALDLYNKYLDQV-IPWOTFDETIK-----ELSRFKQEYSQA 57
Db 1140 ELEAKT--ELEDTLDTAAQQELRSKREQEVNLLKKTLEBEAKTHEAQIQEMRQHSQA 1197
QY 58 ASVLVGDIKTLLMDSQDKYFEAT-----QTVYWCQ-VATOLLAAYILL-----FDEYNEK 107
Db 1198 VEELAEQLE-----QTKRVKANLEKAKOTLENERGELANEV---KVLQKGDSEHKRK 1248
QY 108 KASAO-KDILIKV-----LDGITKLINEAOKSLVSSQSFNNASGKLALD---SQ 154
Db 1249 KVEAQLOELQVFNENGERVTELADKVTKL-----QVELDNVTGLLSQSDSKSK 1298
QY 155 LTNDFSEKSSYPQSQVDKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGK----- 206
Db 1299 LTKDPSALESQDQTQELIQEENRQK-----LPELKNKLKSVQNFPTLLSNTVKQANKDIDAAKLKTTEI 1345
QY 207 -----LIPELKNKLKSVQNFPTLLSNTVKQANKDIDAAKLKTTEI 247
Db 1346 EEEEAHNLKQIATLHAQVADMMKKMDESVGCLETAEEVVKRKLQKDLGLSQRHEEKV 1405
QY 248 AATGKTKTETETTRFVYDDYDMLSLKLEAKKMINTCNEYOKR 291
Db 1406 AAYD--KLEKTKTRLQOELEDDLLVDLDHQOQ-----SACNLEKQ 1443

RESULT 47
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match
Best Local Similarity 6.6%; Score 99.5; DB 4; Length 3878;
Matches 65; Conservative 51; Mismatches 108; Indels 101; Gaps 12;

QY 10 VEVVKNAIETADGALDLYNKYLDQV-IPWOTFDETIKELSRFKQEYSQAASVLVGDIKTL 68
Db 169 LEMSELAGQKQHEIBELNRELEEMRVYGTI--EGLQOL---QEPEAAIKQRDGIITOL 222
QY 69 LMDSQDKYFEATQTVYWCQVGA-----TOLLAAYILLFDEYNEKASAKDILIK 118
Db 223 TANLQARREKDETWREFLETEQSKLIQIQFOQLQASETL-----RNSTHSSAADLL-- 276
QY 119 VLDDGITKLINEAOKSLVSSQSFNNASGKLALDLSQTNDFSEKSSYFQSQV-----DKIR 174
Db 277 -----QAKQOILTHQOQ-----LEEQDHLEEDYQKKKEDFTMQISFLQEKIK 318
```

```
Db 452 YIGPYVTGKNAFPMVN 469
;
; LENGTH: 1960
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P35579
US-09-538-092-1077

Query Match
Best Local Similarity 6.6%; Score 99.5; DB 4; Length 1078;
Matches 70; Conservative 57; Mismatches 110; Indels 105; Gaps 16;

QY 7 DKTVEVVKNAIETADGALDLYNKYLDQVIPWOTFDETIKELSRFKQEYSQAASVLVGDIK 66
Db 609 EETVK-AKEELETLSKID-----NLEKELKQOQSKNE-----LEGOLQ 647
QY 67 TLLMDSQDKYFEATQTVYWCQVATQALLAAYILLFDEY-----NEKASAKDILIKVL- 120
Db 648 NITDSTNEFKE-----LEDELKSIKSNKEISSQNSSELIOKLE 686
QY 121 -----DDGITKLINEAOKSLVSSQSFNNASGKLALDLSQTL-----NDFSEKSSY 165
Db 687 KTEKDLQAKDEEDKLAETKS-----NIDLNSELISLQSKLKEAEESHSSTKDEHSS 740
QY 166 FQSQVDKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKL--IPELKNKLKSVQNFPT 223
Db 741 LSENLKKL-KEEVENTKTSIAKLSAKIEHKKATDEITKTKHTLDQEEHAKOKSOFE 799
QY 224 TLSNTVK---QANKDI--DAAKL-----KLTEIAAIGETK-----ETETT 260
Db 800 SERNDIKSNLDEANKELSDNREKLSNLEKEKTELANNKLKTOBEKISDLETSVAISEDKSK 859
QY 261 RFVYDDYDMLSLKLEAKKMINTCNEYOKRHKGKTLFVPE 302
Db 860 SLKHDIED-----LKREKIKLETTLKE-----NEETMPEKKE 891

RESULT 46
US-09-538-092-1077
; Sequence 1077, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 1077
```

Qy 175 -----KEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNK 214  
Db 319 VYEMQDKVNSNKEBIOKET-----IIIELNTKIIEBEKKTLEKOK 363  
Qy 215 LKSVQNFETLSNTVKQAKNDIDAOKLITTEIAAIGETETTRFYVDYDDMLSL 274  
Db 364 LTTADKLLGELQEQVQKQEIKNKLELT-----NSKQKROSS----- 403  
Qy 275 KEAAKMINTCNEYQKRGKKTLP 299  
Db 404 -EIKQLMGTVLELQKRNKDSQPE 427

## RESULT 48

US-09-447-497-15  
; Sequence 15, Application US/09447497  
; Patent No. 6773911  
; GENERAL INFORMATION:  
; APPLICANT: Penninger, Josef M.  
; APPLICANT: Kroemer, Guido P.  
; APPLICANT: Siderovski, David P.  
; APPLICANT: Zamzami, Naoufal  
; APPLICANT: Sustin, Santos A.  
; APPLICANT: Snow, Bryan E.L.  
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR  
; FILE REFERENCE: 01017/36780  
; CURRENT APPLICATION NUMBER: US/09/447,497  
; CURRENT FILING DATE: 1999-11-23  
; EARLIER APPLICATION NUMBER: 60/109,595  
; EARLIER FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 526  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human apoptosis-inducing factor (hAIF isoform #2);  
; OTHER INFORMATION: hAIF-exon-skip-Gold; mature polypeptide  
US-09-447-497-15

Query Match 6.6%; Score 99; DB 4; Length 526;

Best Local Similarity 21.8%; Pred. No. 0.76; Indels 76; Gaps 13;  
Matches 60; Conservative 46; Mismatches 93;

Qy 37 WQTFDETIKELSRFKQ-----EYSQAAS-----VLVDIKTLMD 72  
Db 93 WFSDDPNVTKTLRFKQWNGKERSIYFQPPSYSAQDLPHIENGVAVLTK-KVQLDV 151  
Qy 73 QD---KYFEATQTVYEWCGVATQLLAAYILLFDEYNEKASQAQ---KDILIKVLDD--GI 124  
Db 152 RDNVKNLNDGSGQIYKELCLATGTGTPRSLAID-----RAGAEVKSRITLFRKIGDPRSL 206  
Qy 125 TKLNEAKSLVSSQSFNNA-----SGKLLALDSQLTNDFSEKSS-----YFQS-QV 170  
Db 207 EKISREVKSTIIGGGFLGSELACALGRKARALGTEVIQLPPEKGNMGKILPEYLSNWTM 266  
Qy 171 DKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNKLSQVQNFETLSNTVK 230  
Db 267 EKVRRE-----GVKVPNAIVQS-----VGVSSGKLLIKLDGRK-----V 302  
Qy 231 QANKDIDAOKLITTEIAAIGETETTRFYVD 265  
Db 303 ETDHVAAGVLEPNVELAKTGGLSDSDFGGFRVN 337

## RESULT 49

US-09-447-497-14  
; Sequence 14, Application US/09447497  
; Patent No. 6773911  
; GENERAL INFORMATION:  
; APPLICANT: Penninger, Josef M.

; APPLICANT: Kroemer, Guido P.  
; APPLICANT: Siderovski, David P.  
; APPLICANT: Zamzami, Naoufal  
; APPLICANT: Sustin, Santos A.  
; APPLICANT: Snow, Bryan E.L.  
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR  
; CURRENT APPLICATION NUMBER: US/09/447,497  
; CURRENT FILING DATE: 1999-11-23  
; EARLIER APPLICATION NUMBER: 60/109,595  
; EARLIER FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 553  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-447-497-14

Query Match 6.6%; Score 99; DB 4; Length 553;

Best Local Similarity 21.8%; Pred. No. 0.82; Indels 76; Gaps 13;  
Matches 60; Conservative 46; Mismatches 93;

Qy 37 WQTFDETIKELSRFKQ-----EYSQAAS-----VLVDIKTLMD 72  
Db 120 WFSDDPNVTKTLRFKQWNGKERSIYFQPPSYSAQDLPHIENGVAVLTK-KVQLDV 178  
Qy 73 QD---KYFEATQTVYEWCGVATQLLAAYILLFDEYNEKASQAQ---KDILIKVLDD--GI 124  
Db 179 RDNVKNLNDGSGQIYKELCLATGTGTPRSLAID-----RAGAEVKSRITLFRKIGDPRSL 233  
Qy 125 TKLNEAKSLVSSQSFNNA-----SGKLLALDSQLTNDFSEKSS-----YFQS-QV 170  
Db 234 EKISREVKSTIIGGGFLGSELACALGRKARALGTEVIQLPPEKGNMGKILPEYLSNWTM 293  
Qy 171 DKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNKLSQVQNFETLSNTVK 230  
Db 294 EKVRRE-----GVKVPNAIVQS-----VGVSSGKLLIKLDGRK-----V 329  
Qy 231 QANKDIDAOKLITTEIAAIGETETTRFYVD 265  
Db 330 ETDHVAAGVLEPNVELAKTGGLSDSDFGGFRVN 364

## RESULT 50

US-09-447-497-12  
; Sequence 12, Application US/09447497  
; Patent No. 6773911  
; GENERAL INFORMATION:

; APPLICANT: Penninger, Josef M.  
; APPLICANT: Kroemer, Guido P.  
; APPLICANT: Siderovski, David P.  
; APPLICANT: Zamzami, Naoufal  
; APPLICANT: Sustin, Santos A.  
; APPLICANT: Snow, Bryan E.L.  
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR  
; FILE REFERENCE: 01017/36780  
; CURRENT APPLICATION NUMBER: US/09/447,497  
; CURRENT FILING DATE: 1999-11-23  
; EARLIER APPLICATION NUMBER: 60/109,595  
; EARLIER FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 609  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human apoptosis-inducing factor (hAIF isoform #1)  
; OTHER INFORMATION: hAIF-alt-exon-Gold; mature polypeptide  
US-09-447-497-12

Query Match 6.6%; Score 99; DB 4; Length 609;



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:52:17 ; Search time 17.4424 Seconds  
(without alignments)  
1671.423 Million cell updates/sec

Title: US-09-993-292B-28  
Perfect score: 1508  
Sequence: 1 MTEIVADKTVEVVKNAIETA.....TCNEYQKSHCKKTLFVPEV 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : PIR 79:\*\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description           |
|------------|-------|-------------|--------|----------|-----------------------|
| 1          | 1508  | 100.0       | 305    | 2 C64864 | hemolysin E - Esch    |
| 2          | 1493  | 99.0        | 305    | 2 E90838 | hemolysin E [impor    |
| 3          | 1493  | 99.0        | 305    | 2 E85696 | probable pore form    |
| 4          | 1400  | 92.8        | 305    | 2 AE0673 | haemolysin HlyE [i    |
| 5          | 133.5 | 8.9         | 584    | 2 S75986 | hypothetical prote    |
| 6          | 126   | 8.4         | 652    | 2 B59102 | hypothetical prote    |
| 7          | 122   | 8.1         | 1496   | 2 T05634 | hypothetical prote    |
| 8          | 120   | 8.0         | 1039   | 2 S18139 | myosin heavy chain    |
| 9          | 117.5 | 7.8         | 927    | 2 AH1369 | transmembrane prot    |
| 10         | 117.5 | 7.8         | 1098   | 2 B70232 | hypothetical prote    |
| 11         | 116.5 | 7.7         | 927    | 2 AG1739 | transmembrane prot    |
| 12         | 115.5 | 7.7         | 664    | 2 A97222 | membrane associate    |
| 13         | 115   | 7.6         | 1127   | 2 T28317 | ORF MSV156 hypothe    |
| 14         | 115   | 7.6         | 1964   | 2 A59282 | nonmuscle myosin I    |
| 15         | 115   | 7.6         | 2017   | 1 A36014 | myosin II heavy chain |
| 16         | 115   | 7.6         | 2057   | 2 S61477 | myosin II heavy ch    |
| 17         | 114.5 | 7.6         | 1269   | 2 E84730 | probable myosin he    |
| 18         | 114   | 7.6         | 796    | 2 B84800 | probable alpha-car    |
| 19         | 114   | 7.6         | 1738   | 2 T14867 | interactin - slime    |
| 20         | 114   | 7.6         | 2401   | 2 T28676 | rhothry protein -     |
| 21         | 113.5 | 7.5         | 1024   | 2 S10056 | hemolysin A - Esch    |
| 22         | 113   | 7.5         | 622    | 2 T22716 | hypothetical prote    |
| 23         | 113   | 7.5         | 955    | 1 A3524  | leukotoxin A - Pas    |
| 24         | 112.5 | 7.5         | 1679   | 2 S48385 | hypothetical prote    |
| 25         | 112   | 7.4         | 295    | 2 G97827 | hypothetical prote    |
| 26         | 112   | 7.4         | 631    | 2 JC4298 | hyaluronan recepto    |
| 27         | 112   | 7.4         | 1875   | 2 S38173 | myosin-like protei    |
| 28         | 111.5 | 7.4         | 821    | 2 S67087 | hypothetical prote    |
| 29         | 111.5 | 7.4         | 2139   | 2 T18296 | myosin heavy chain    |

ALIGNMENTS

RESULT 1

C64864

hemolysin E - Escherichia coli (strain K-12)

N;Alternate names: hemolysin-inducing protein

C;Species: Escherichia coli

C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C;Accession: C64864

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: C64864

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-305 <BLAT>

A;Cross-references: GB:AE000216; GB:U00096; NID:G1787417; PIDN:AAC74266.1; PID:G1787430;

A;Experimental source: strain K-12, substrain MGL55

C;Genetics:

A;Gene: hlyE; hpr

C;Function:

A;Description: hemolytic activity

A;Note: pore formation

C;Superfamily: Escherichia coli hemolysin E

C;Keywords: cytolysis; cytotoxin; hemolysis; transmembrane protein

F;181-197/Domain: transmembrane #status predicted <TM>

F;123/Active site: Asp #status predicted

Query Match 100.0%; Score 1508; DB 2; Length 305;  
Best Local Similarity 100.0%; Pred. No. 2.4e-87;  
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPQTFDETIKLSRFKQEYSQAASV 60

Db 3 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPQTFDETIKLSRFKQEYSQAASV 62

Qy 61 LVGDITKLMDSQDKYFEATQTVYVCGVATQLLAAYILLFDYNEKKAQAQDKILIKVL 120

Db 63 LVGDITKLMDSQDKYFEATQTVYVCGVATQLLAAYILLFDYNEKKAQAQDKILIKVL 122

Qy 121 DDGITKLNEAKSLLYSQSFNNASGKLLALDLSQLTNDSEKSYFQSOVDKIRKEAYAG 180

Db 123 DDGITKLNEAKSLLYSQSFNNASGKLLALDLSQLTNDSEKSYFQSOVDKIRKEAYAG 182

Qy 181 AAGVAGPPGLIISYSIAAGVVEGKLIPELKNKLSQVQNFVFTLSNTVKQANKDIDAAK 240

Db 183 AAGVAGPPGLIISYSIAAGVVEGKLIPELKNKLSQVQNFVFTLSNTVKQANKDIDAAK 242

Qy 181 AAGVAGPPGLIISYSIAAGVVEGKLIPELKNKLSQVQNFVFTLSNTVKQANKDIDAAK 240

Db 183 AAGVAGPPGLIISYSIAAGVVEGKLIPELKNKLSQVQNFVFTLSNTVKQANKDIDAAK 242











```
Query Match 7.6%; Score 115; DB 2; Length 1127;
Best Local Similarity 20.2%; Pred. No. 18;
Matches 70; Conservative 51; Mismatches 124; Indels 102; Gaps 11;

QY 5 VADKTEV--VKNAETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQEYSQAASVLV 62
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
192 INNKEIEFNIDNVOKEINKODELKNL-----DESKKEFKKQEEELNKTIDKQK 242
QY 63 GDIKTLMDSQ----DKYFEATQTVYEWCGVATQLAAAYILLDFBYNEKKASQAQDILI 117
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
243 EELIKKLNDEKINFINDEKQKLLDQINSKINTLNENIKGMVNLNTEKNIKSLNQLNEILN 302
QY 118 KVLDDGHTKLENAQKSLVSSQSFNNASGKLLALDSQTN-----DFSEKSSYFQ 167
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
303 K--DSTIKSLDEKQKLLDELKNINITSYLNKSNITKITNIQQLLESSLTFDNNANIN-- 358
QY 168 SQVDKIRKEAYAGAAAGVAGPFGILLISYSIAAGVVEGKLIPELKNKLSQVQVFFTTLSN 227
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
359 -----INELSKIKLFDNDIQKLN 378
QY 228 TVKQANKDID-----AAKLKLTTHIAAIGEIKTET-----ETTRFY----- 263
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
379 DITEQNNKITDFNNSTRIPEKLDTEYKIDDIKNNNLQKLESYKKIDEOTEYKNNKI 438
QY 264 -VDYDLM-----LSLLKEAAKMMINTCNEYQKRGHKTKTLFEVPEV 303
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
439 NKEYNDIIEELKNNNLQKLEENKK-IDQTEYKNNKINKEYNDIIEEL 484

RESULT 14
A59282
nonmuscle myosin II heavy chain A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C:Accession: A59282
R:Phai.Dey, N.; Taira, M.; Conti, M.A.; Nooruddin, H.; Adelstein, R.S.
Mech. Dev. 78, 33-36, 1998
A:Title: Differential expression of non-muscle myosin heavy chain genes during Xenopus
A:Reference number: A59282; MUID:199077683; PMID:9858676
A:Accession: A59282
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1964 <BA>
A:Cross-references: UNIPROT:Q93522; GB:AF055895; NID:G3660671; PIDN:AAC83556.1; PID:G366
A:Experimental source: cell line XTC
C:Superfamily: myosin heavy chain; myosin motor domain homology
F;84-764/Domain: myosin motor domain homology <MMO>

Query Match 7.6%; Score 115; DB 2; Length 1964;
Best Local Similarity 18.8%; Pred. No. 36;
Matches 60; Conservative 64; Mismatches 113; Indels 82; Gaps 13;

QY 8 KTEVVKNAETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQEYSQAASVLVAG---- 63
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1174 KTLK-----DEARTHEGQTEIRQKHSQAV--EELSEQLQTKRLKGNLEKAKQALGERNE 1228
QY 64 ---DIKTLMDSQDYFEATQTVYEWCGVATQLAAAYILLDFBYNEKKASQAQ-KDILIKV 119
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1229 LANEVKTLLQKGD-----SEHKRKVEAQLQELQVKV 1261
QY 120 LD-DGI--TKLNEAKSLVSSQSFNNASGKLLALDSQTNDFSEKSSYFQSDVKIRKEA 177
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1262 TEGDRVRSSELSEKANRLQVELDNNVLSLSDSKSLKLGKDFSTLESQFQDAQLQEST 1321
QY 178 YAGAAAGVAGPFGILLISYSIAAGVVEGK--LIPEL-----KNKLSQVQVFFTTLSN 227
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1322 RQK-----LSFSTKLQMEDKNGNLEQLLEEEBAKKNLCKQISTLQSQMTD 1368
QY 228 TVKQANKD-----IDAAKLKLTTHIAAIGE-----IKTETTRFYVDYDMLLS 272
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1369 MKKXNDVNSLETVEELKKLQKDLQKLEAVNORFEKAAAYDKLEKTKTRLQQLDLDISVD 1428
```

```
QY 273 LLKEAAKMMINTCNEYQKR 291
DB 1429 L--DHQRQIVSNLEKKQKK 1445

RESULT 15
A36014
myosin heavy chain, nonmuscle - fruit fly (Drosophila melanogaster)
N:Contains: myosin ATPase (BC 3.6.4.1)
C:Species: Drosophila melanogaster
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
R:Ketchum, A.S.; Stewart, C.T.; Stewart, M.; Kiehart, D.P.
Proc. Natl. Acad. Sci. U.S.A. 87, 6316-6320, 1990
A:Title: Complete sequence of the Drosophila nonmuscle myosin heavy-chain transcript: con
A:Reference number: A36014; MUID:90349606; PMID:2117279
A:Accession: A36014
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2017 <KET>
A:Cross-references: UNIPROT:Q99323; GB:M35012
C:Genetics:
A:Gene: FlyBase:zip
A:Cross-references: FlyBase:FBgn0005634
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F;1-2017/Product: myosin heavy chain, form I #status predicted <MA1>
F;16-2017/Product: myosin heavy chain, form II #status predicted <MA2>
F;135-815/Domain: myosin motor domain homology <MMOT>
F;225-232/Region: nucleotide-binding motif A (P-loop)
F;598-631/Region: actin binding #status predicted
F;705-727/Region: actin binding #status predicted
F;888-2017/Domain: coiled coil #status predicted <COI>
F;888-1328/Region: S2
F;1329-2017/Region: light meromyosin
F;176/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F;231/Binding site: ATP (Lys) #status predicted
F;745,755/Active site: Cys #status predicted

Query Match 7.6%; Score 115; DB 1; Length 2017;
Best Local Similarity 21.3%; Pred. No. 37;
Matches 80; Conservative 48; Mismatches 107; Indels 140; Gaps 16;

QY 14 KNAIETADGALDLYNKYL-----DQVTPWQTFDETIKELSRFK 51
DB 1263 KTVLEKAKGTLEAENADLATELSVNSRQENRRRKRQAEISQIAELQV---KLAETEAR 1319
QY 52 QEYSQAASVLVGDITLLMDSQDYFEATQTVYEWCGVATQLAAAYILLDFBYNEK--- 107
DB 1320 SELQEKCTKLQEAENITNQLAEALKASAAVKSASNMESQLTEAQLLEETRQKLGSL 1379
QY 108 -----KASAQKDI-LIKVLDDGIT 125
DB 1380 SKLRQISEKALQEQLEEDDEAKRYERKLAETVTTQMQLKKAEEDADLAKLESGKK 1439
QY 126 KLN-----BAQ-KSLIVSSQSFNNASGKL-----LALDSQTNDFP-EKSSYFQ 169
DB 1440 RLNKDIEALERQVKELAQNDRLDKSKKIQSELEADTIEAQRKTVLEKK---QKN 1496
QY 170 VDKTRKAYAGAAAGVAGPFGILLISYIA-----AGVVEGKL----- 207
DB 1497 FDKILAEKA-----ISEQIAQRDTAREAREAREKETKVLVSRELDEAFDK 1542
QY 208 IPELKNKLSQVQVFFTTLSNTVQANK---DIDAAKLKLTTHIAAIGEITETTRFYV 264
DB 1543 IEDLENKTKTLQNELDDLANTQGTADKNVHELEKAKALESQLA---ELKAQNEELE--- 1596
QY 265 DYDDLMLSLLEAKK 279
DB 1597 --DDLQLT---EDAK 1606

RESULT 16
```

S61477  
myosin II heavy chain, non-muscle - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C>Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004  
C/Accession: S61477; S65349  
R:Mansfield, S.G.; Al-Shirawi, D.Y.; Ketchum, A.S.; Newbern, E.C.; Kiehart, D.P.  
J. Mol. Biol. 255, 98-109, 1996  
A/Title: Molecular organization and alternative splicing in zipper, the gene that encodes myosin II heavy chain  
A/Reference number: S61477; MUID:96144835; PMID:8568878  
A/Accession: S61477  
A/Molecule type: DNA  
A/Residues: 1-2057 <MAN>  
A/Cross-references: UNIPROT:Q94987; EMBL:U35816  
R:Mansfield, S.G.; Al-Shirawi, D.Y.; Ketchum, A.S.; Newbern, E.C.; Kiehart, D.P.  
submitted to the EMBL Data Library, September 1995  
A/Reference number: S65349  
A/Accession: S65349  
A/Molecule type: DNA  
A/Residues: 1-1908, 'NL', 1911-2057 <MAN>  
A/Cross-references: EMBL:U35816; NID:g1141789; PIDN:AAB09049.1; PID:g1572481  
C/Genetics: zip  
A/Cross-references: FlyBase:FBgn0005634  
A/Introns: 10/3; 90/3; 162/3; 215/1; 255/3; 263/3; 303/3; 334/3; 1248/1; 1349/3; 1526/3;  
C/Superfamily: myosin heavy chain; myosin motor domain homology  
C/Keywords: alternative splicing; ATP; nucleotide binding; P-loop  
F:135-855/domain: myosin motor domain homology <MOT>  
F:225-232/Region: nucleotide-binding motif A (P-loop)  
  
Query Match 7.6%; Score 115; DB 2; Length 2057;  
Best Local Similarity 21.3%; Pred. No. 38;  
Matches 80; Conservative 48; Mismatches 107; Indels 140; Gaps 16;  
  
Qy 14 KNAIETADGALDLYNKKYL-----DQVTPWQTFDETTELKSRFK 51  
Db KTVLEKAKGTLEAENADLATELRVNSRQENRRRKAESQIAELQV---KLAIEIRAR 1359  
  
Qy 52 QEYSQAASVLVDIKTLMSQDKYFEATQTVVWCGVATQLLAAAYILLDFEYNEK----- 107  
Db SELQEKCTKQQAENITNQLAEALKASAAVKSANMESQLTEAQQLLSEETRQKLGSL 1419  
  
Qy 108 -----KASAKQDI-LIKVLDDGIT 125  
Db SKLRQISEKALQEOLEEDDEAKRNVKRLAEVTTQWQETKKAEDADLAKLEGGK 1479  
  
Qy 126 KLN-----EAQ-KSLIVSQSFNNSGKL-----LALDSQLTNDFS-EKSSYFQSQ 169  
Db RLKNDIEALERQVKELIAQNDRDLKSKKIQSELEDATELEAQRKTVLELEK---QKN 1536  
  
Qy 170 VDKIRKAYAGAAVGVAGPGLIISYIA-----AGVVEGKL----- 207  
Db FDKILAEKA-----ISEQIAQERDTAREAREKETKVLVSRLDEAFDK 1582  
  
Qy 208 IPELKNLKSQNFPTLSTNTVKQANK---DIDAAKLTTETIAAIGETKTETTRFYV 264  
Db IEDLENKRTKLQNELDLANTQGTADKNVHELEKAKALSSQLA---ELKQNEELE--- 1636  
  
Qy 265 DYDDLMLSLKKAQK 279  
Db 1637 --DDLQLT---EDAK 1646  
  
RESULT 17  
F84730  
probable myosin heavy chain [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C/Accession: F84730  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: F84730  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1269 <STO>  
A/Cross-references: GB:AE002093; NID:g6598483; PIDN:AAC69932.2; GSPDB:GN00139  
C/Genetics:  
A/Gene: At2g32240  
A/Map position: 2  
  
Query Match 7.6%; Score 114.5; DB 2; Length 1269;  
Best Local Similarity 18.0%; Pred. No. 22;  
Matches 56; Conservative 63; Mismatches 123; Indels 69; Gaps 9;  
  
Qy 1 MTEIVAD--KTVEVVKNAIETADGALDLYNKKYLDQVTPWQTFDETTELKSRFKQYEQAA 58  
Db LTKLRDLGKIKSYERQALAEASGSSSLKLEQLTGLRLAAASVNE--KLQEFQDAQ 882  
  
Qy 59 SVLVGDIKTLMSQDKYFEATQTVVWCGVATQLLAAAYILLDFEYNEKKAQKDIILK 118  
Db KSSQSSSELLAETNNQLK---IKIQLEGLI-----GGSGVKEKTAALK 925  
  
Qy 119 VLDDGITKNEAQKSLIVSQSFNNSGKLALDSQLTNDFSQSYFQSQVDKIRKEAY 178  
Db RLEEAIERFNQKETE-----SSDLVEKLKTHENQIEYKKLALH 963  
  
Qy 179 AGAAAGVAGPGLIISYIAAGVVEGKLPELKNLKSQNFPTLSTNT-----VKQA 232  
Db -----ASGVADTRKY-ELEDALSKLNLESTTEELGAKCQGLEKE 1003  
  
Qy 233 NKDIDAAKLTTETIAAIGETKTETTRFYVDYDDLMLSLKKAQKMTNCEYQKRH 292  
Db SGDLAEVNLKLNLELANHGSEANELQTKLSALEAKEQTANELEASKTTIEDLTQKLTSE 1063  
  
Qy 293 GKTKLFEVPEV 303  
Db GEKLQSQIEKL 1074  
  
RESULT 18  
B84800  
probable alpha-carboxyltransferase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C/Accession: B84800  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: B84800  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-796 <STO>  
A/Cross-references: UNIPROT:Q7XJRI; GB:AE002093; NID:g4895181; PIDN:AAD32768.1; GSPDB:GN00139  
C/Genetics:  
A/Gene: At2g38040  
A/Map position: 2  
  
Query Match 7.6%; Score 114; DB 2; Length 796;  
Best Local Similarity 24.9%; Pred. No. 13;  
Matches 68; Conservative 44; Mismatches 119; Indels 42; Gaps 10;  
  
Qy 42 ETIKELSRFKQYEQAAASVLVGDIKTLMSQDKYFEATQTVVWCGVATQLLAAAYILLF 101  
Db EQILKAKETSTEAEPSEVLNEMTEKLKSIDDEYTEATAV---GLEERLTA---MR 522  
  
Qy 102 DYNEKKAQKDIILKVLDDGITKNEAQKSLIVSQSFNNSGKLAL-----DS 153  
Db -----KASSEEHLMHVLTEKIEKLEKEFNTRLTDPAPNYESLKSCLNMLRDRSRAKAS 580  
  
Qy 154 QLTFNDFSEKSSYFQSQVDK-----IRKEAYAGAAAGVAGPGLIISYIAAGVVE 204

Mol. Biochem. Parasitol. 42, 241-246, 1990  
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co  
A:Reference number: A45521; MUID:91101660; PMID:2270106  
A:Accession: A45521  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 2260-2401 <KEE>  
A:Cross-references: GB:M34281

Query Match 7.6%; Score 114; DB 2; Length 2401;  
Best Local Similarity 19.4%; Pred. No. 54;  
Matches 80; Conservative 57; Mismatches 141; Indels 134; Gaps 16;

QY 5 VADKTVE---VVKNAIETADGALDLYNKYLDQVTPWQTFDET--IKELSRKQEFY----S 55  
DB 1760 IEDKIIENGLINKLIETKDCMLFYTKLTVELTKIKTDTYKFTSATSKFSKFLKYID 1819  
QY 56 QAASVLVGDITKLLMDSOOKY-----FEATQTVYEWCG 88  
DB 1820 ATSNLSNDDINTL-----QTKYDLNQLNHVASMVADATNDNNNLLIEKEKEATKINN--- 1872  
QY 89 VATQLLA-----AVILLFDEYNEKKAQKIDILIKVLDGDTKLNEAOKS 133  
DB 1873 -LTELFTDSNKIDADGLHNNKIQIIYFNSLHKSIDSITKOLYKQKHAFLKLNIGHINKK 1931  
QY 134 LLVSQSFPNAGKLLALDLSQLTNDFSEKSSYFQSDVKIRKEAYAGAAAGVVGPFGL- 192  
DB 1932 YFDISKEDFN---LQLOESELNLDLKEIGQKISDK--KKQFLHALSETPIPNFNTL 1986  
QY 193 -----IISYSTAAGVVEG-----KLPELKNKLKSVONFPTTL---SNTVK 230  
DB 1987 KEIYHDIVKYKQIDEIENITNEENITLYMDIITKLMKVQSILNFVTVYENDSNIIK 2046  
QY 231 ---QANKDIDAAKLKLTTT-----IAAIGEIKT----- 255  
DB 2047 QHIDNNNDVSKIESLETTSQFQKILNKLNEIKAAQFYDNNNNINNVISTISQDVNDVK 2106  
QY 256 -----ETETRFYVDYDDLMLS---LLKEAAKMTINTCNEYOKRGKK 295  
DB 2107 KHISKDLTIENELIQKSLIEDIKKSTYDIRSEQITKYVNPIDHYVEQOTKK 2158

RESULT 21  
S10056  
hemolysin A - Escherichia coli plasmid pHLy152  
C:Species: Escherichia coli  
C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 04-Mar-2000  
C:Accession: S10056  
R:Hees, J.; Wels, W.; Vogel, M.; Goebel, W.  
FEMS Microbiol. Lett. 34, 1-11, 1986  
A:Title: Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparis  
A:Reference number: S07209  
A:Accession: S10056  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1024 <HES>  
A:Cross-references: EMBL:M14107  
C:Genetics:  
A:Genome: plasmid pHLy152  
C:Superfamily: hemolysin A; hemolysin A homology  
C:Keywords: lipoprotein  
F:247-792/Domain: hemolysin A homology <HLyA>  
F:564,690/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.5%; Score 113.5; DB 2; Length 1024;  
Best Local Similarity 21.4%; Pred. No. 20;  
Matches 70; Conservative 59; Mismatches 121; Indels 77; Gaps 14;

QY 2 TEIVADKTVEYVK-----NAJETADGALDLYNKYLDQVTPWQTFDEIKELSRKQEF 53  
DB 155 TALSSMKIDELIKQKSGNVSSSELAKASIELINQLVDTVASLN-----NNVNSFSQQ 208  
QY 54 YQAASVL-----VGDITKLLMDSQDKYFEATQTVYEWCGVATQLLAAYILL-FDEY 104

Mol. Biochem. Parasitol. 42, 241-246, 1990  
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co  
A:Reference number: A45521; MUID:91101660; PMID:2270106  
A:Accession: A45521  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 2260-2401 <KEE>  
A:Cross-references: GB:M34281

Query Match 7.6%; Score 114; DB 2; Length 2401;  
Best Local Similarity 19.4%; Pred. No. 54;  
Matches 80; Conservative 57; Mismatches 141; Indels 134; Gaps 16;

QY 5 VADKTVE---VVKNAIETADGALDLYNKYLDQVTPWQTFDET--IKELSRKQEFY----S 55  
DB 1760 IEDKIIENGLINKLIETKDCMLFYTKLTVELTKIKTDTYKFTSATSKFSKFLKYID 1819  
QY 56 QAASVLVGDITKLLMDSOOKY-----FEATQTVYEWCG 88  
DB 1820 ATSNLSNDDINTL-----QTKYDLNQLNHVASMVADATNDNNNLLIEKEKEATKINN--- 1872  
QY 89 VATQLLA-----AVILLFDEYNEKKAQKIDILIKVLDGDTKLNEAOKS 133  
DB 1873 -LTELFTDSNKIDADGLHNNKIQIIYFNSLHKSIDSITKOLYKQKHAFLKLNIGHINKK 1931  
QY 134 LLVSQSFPNAGKLLALDLSQLTNDFSEKSSYFQSDVKIRKEAYAGAAAGVVGPFGL- 192  
DB 1932 YFDISKEDFN---LQLOESELNLDLKEIGQKISDK--KKQFLHALSETPIPNFNTL 1986  
QY 193 -----IISYSTAAGVVEG-----KLPELKNKLKSVONFPTTL---SNTVK 230  
DB 1987 KEIYHDIVKYKQIDEIENITNEENITLYMDIITKLMKVQSILNFVTVYENDSNIIK 2046  
QY 231 ---QANKDIDAAKLKLTTT-----IAAIGEIKT----- 255  
DB 2047 QHIDNNNDVSKIESLETTSQFQKILNKLNEIKAAQFYDNNNNINNVISTISQDVNDVK 2106  
QY 256 -----ETETRFYVDYDDLMLS---LLKEAAKMTINTCNEYOKRGKK 295  
DB 2107 KHISKDLTIENELIQKSLIEDIKKSTYDIRSEQITKYVNPIDHYVEQOTKK 2158

RESULT 21  
S10056  
hemolysin A - Escherichia coli plasmid pHLy152  
C:Species: Escherichia coli  
C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 04-Mar-2000  
C:Accession: S10056  
R:Hees, J.; Wels, W.; Vogel, M.; Goebel, W.  
FEMS Microbiol. Lett. 34, 1-11, 1986  
A:Title: Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparis  
A:Reference number: S07209  
A:Accession: S10056  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1024 <HES>  
A:Cross-references: EMBL:M14107  
C:Genetics:  
A:Genome: plasmid pHLy152  
C:Superfamily: hemolysin A; hemolysin A homology  
C:Keywords: lipoprotein  
F:247-792/Domain: hemolysin A homology <HLyA>  
F:564,690/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.5%; Score 113.5; DB 2; Length 1024;  
Best Local Similarity 21.4%; Pred. No. 20;  
Matches 70; Conservative 59; Mismatches 121; Indels 77; Gaps 14;

QY 2 TEIVADKTVEYVK-----NAJETADGALDLYNKYLDQVTPWQTFDEIKELSRKQEF 53  
DB 155 TALSSMKIDELIKQKSGNVSSSELAKASIELINQLVDTVASLN-----NNVNSFSQQ 208  
QY 54 YQAASVL-----VGDITKLLMDSQDKYFEATQTVYEWCGVATQLLAAYILL-FDEY 104

Db 209 LNTGSLNNTKHLNGVN-KLQNPNDNIGAGLDTV---SGLSAISASFILSNADAD 264  
Qy 105 NEKASAKOILIKVLDD---GITKLNEAQSLVSSQSFNNASGKLLALDSQTLNDFSE 161  
Db 265 TRTKAAGVELTTKVLNGVNGISQYIAQR---AAQGLSTSAAGLASAVTLAISP 320  
Qy 162 KSSVFQSOVDKIRK----EYAGAAAGVAVGPGFLIISYSIAAGVVGKLIPELKNKLKS 217  
Db 321 LS--FLSIADKPKRANKIEYSQRPKLGVGDGDSLLAAPHKETGAIDASL----- 368  
Qy 218 VQNFPTTLNNTVKQANKIDDAKLLKLTETI-----AAIGEIKTETETTFYVDYD 267  
Db 369 -----TTISTVLASVSGISNA---ATSLVGAPVSLVGNVGTGIIISGLEASK----- 414  
Qy 268 DLMLSLKEAKAKMINTCNEYQKRHK 294  
Db 415 ---QAMFEHVASKWADVIAEWKKGK 438

RESULT 22  
T22716  
hypothetical protein F55C5.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T22716  
R:Harris, B.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z19604  
A:Accession: T22716  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-622 <WIL>  
A:Cross-references: UNIPROT:Q20822; EMBL:Z78198; PIDN:Z78198; GSPDB:GN00023; CESP:F55C5  
A:Experimental source: clone F55C5  
C:Genetics:  
A:Gene: CESP:F55C5.8  
A:Map position: 5  
A:Introns: 24/3; 44/2; 147/3; 251/3; 561/3

Query Match 7.5%; Score 113; DB 2; Length 622;  
Best Local Similarity 20.6%; Pred. No. 11;  
Matches 67; Conservative 54; Mismatches 100; Indels 104; Gaps 15;

Qy 41 DETIKELSRFKQYSQAASVLGVGDIKTLMDSDQKFEATQTVYEW-----CGVAT 91  
Db 251 DKLISEM-RASATSAEVVTIEWGAKSTVDDE----KAKQVQEWKQTEVELAQCCQTPK 304  
Qy 92 QLLAAAYILLDFEYNEKASAKOILIKVLDDGITYKLE-AQKSLVSSQSF-----NNA 144  
Db 305 EKWALP-----EKATADTRDAIDRISDIIRKSSSENADTTVLQSIKAYLEFLKNGT 356  
Qy 145 SGKLLALDSQTLNDFSEKSSYFQS-----QVDKIRKE 176  
Db 357 ASRYLAI---IDNTKSEKSKFPQDLLRLYDSVIEIYKEVABIPGADHKNLIQAFVKVE 413  
Qy 177 AYAGAAAGVAVGPGFLIISYSIAAGV-----VEGKLIPELKNKLKS-----VQ 219  
Db 414 YYRAFRCFYMASSYSALHKYSEAAALPDRTVRSRVQDAEGKL-----KCLKSSSFITNETQ 468  
Qy 220 NFFTTLNNTVKQANKIDDAKLLKLTETIAGETITETTFYVDYDMLSLKKA 279  
Db 469 SSLNELRSEVESAKVTVRAARL-----ASAAGDKVTDSELAKE-ID-----KR 510  
Qy 280 KMINTCNEYQK---RHGKKTILFEVP 301  
Db 511 PLLETVNEWRQWVRNLSLKOKKTIIP 535

RESULT 23  
A35254  
leukotoxin A - Pasteurella haemolytica (serotype T10)  
N:Alternate names: lktA protein

C:Species: Pasteurella haemolytica  
C:Date: 10-Aug-1990 #sequence\_revision 15-Nov-1996 #text\_change 09-Jul-2004  
C:Accession: S37145; A35254; S34237; S34235  
R:Lainson, A.F.; Aitchison, K.; Donachie, W.  
submitted to the EMBL Data Library, September 1993  
A:Description: DNA sequence of the leukotoxin A gene from P. haemolytica T10 serotype.  
A:Reference number: S37145  
A:Accession: S37145  
A:Molecule type: DNA  
A:Residues: 1-955 <LA1>  
A:Cross-references: UNIPROT:P55117; EMBL:Z26247; NID:G400424; PIDN:CAA81206.1; PID:G4004  
R:Highlander, S.K.; Engler, M.J.; Weinstock, G.M.  
J. Bacteriol. 172, 2343-2350, 1990  
A:Title: Secretion and expression of the Pasteurella haemolytica leukotoxin.  
A:Reference number: A35254; MUID:90236888; PMID:2185213  
A:Accession: A35254  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 950-955 <HIG>  
A:Cross-references: GB:M24197; GB:M34943; GB:M34944  
R:Lainson, A.F.; Aitchison, K.D.; Donachie, W.  
submitted to the EMBL Data Library, June 1993  
A:Description: DNA sequence of the carboxy terminal end of leukotoxin A from the T3 sero  
A:Reference number: S34235  
A:Accession: S34237  
A:Molecule type: DNA  
A:Residues: 745-955 <LA2>  
A:Cross-references: EMBL:Z22884; NID:G311828; PIDN:CAA80498.1; PID:G311829  
A:Experimental source: serotype T3  
A:Accession: S34235  
A:Molecule type: DNA  
A:Residues: 723-955 <LA3>  
A:Cross-references: EMBL:Z22887; NID:G311824; PIDN:CAA80501.1; PID:G311825  
A:Experimental source: serotype T10  
C:Function:  
A:Description: attacks cell membranes and causes cell lysis  
C:Superfamily: hemolysin A; hemolysin A homology  
C:Keywords: calcium binding; cytolysis; exotoxin; hemolysis; lipoprotein; tandem repeat;  
F:240-786/Domain: hemolysin A homology <HLXA>  
F:718-809/Region: 9-residue repeats (G-G-X-G-[DN]-D-X-[LVIYF]-X)  
F:718-726/Region: repeat  
F:727-735/Region: repeat  
F:736-744/Region: repeat  
F:745-753/Region: repeat  
F:754-762/Region: repeat  
F:763-771/Region: repeat  
F:772-780/Region: repeat  
F:781-789/Region: repeat  
F:792-800/Region: repeat  
F:801-809/Region: repeat  
F:556/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.5%; Score 113; DB 1; Length 955;  
Best Local Similarity 18.6%; Pred. No. 19;  
Matches 69; Conservative 67; Mismatches 136; Indels 98; Gaps 12;

Qy 6 ADKTEVVEVNAIETADGALDLYNKYLDQVTPWOTFDETIKELSRFKQYSQAASVLVG-- 63  
Db 89 AQTSLGTIQNVLGITRGIVLSAPQLDKLQKNKGQALGSSESTAQNFSQAKTVLSGVQ 148  
Qy 64 -----DIKTLMDSDQK-----FEATQTVYEWGVATQQLAAAYILLDFEYNEK 107  
Db 149 GNSRTVLAGMDLDEALQNESDQLTLAKAGLELTNLSLTIANSVQTLDAFSEQISQFGSK 208  
Qy 108 KASAQKDILIKVLDD---GITKLNEAQSLVSSQSFNNASGKLLALD----- 152  
Db 209 LQNVKG---LGAAGDKLKNIGGLDKAGLGDVKSRLLSGATAALVLADKDASTAKKVAG 265  
Qy 153 ----SQTNDFFSEK-SSYFQSOVDKIRKEAYAGAAAGVAVGPGFLIISYSIAAGV--VE 204  
Db 266 FELANQVGNITRAVSSYILAQ-----RVAAGLSTSGPVAALIASVVAISPLS 315  
Qy 205 GKLIPELKNKLKSVQNF-----TTLSNTVKQANKIDDAKLLKLTETIAAI----- 250

```

Query Match          7.4%; Score 112; DB 2; Length 295;
Best Local Similarity 24.9%; Pred. No. 5;
Matches 56; Conservative 40; Mismatches 83; Indels 46; Gaps 11;

QY      94 LAAYILLPDEYNEKASQKDLIKVLDDGITKLE--AQSKLLVSSQSFNNASGKLIAL 151
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      4 LLLIITVFTFNVAQASL-PNIVASVNDPIT-LNEFPARKMIM---ALNNVESLTPAQ 58
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY     152 DSQLN---DFSEKSSVFQGVOK-IRKEAYAGAAAGV---AGPFLGIISYIAAGV 203
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB     59 DQKLSDLALKSLIDSLFLPQAGREIPQERIEAIAKSIEDRNKMPHGSLLQYLKSRVN 118
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY     204 EGKLPKLNKILKSQNFPTTILSTVTKQAKNDIDAAKLKLTETAAIGEIKTETTRFY 263
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB     119 PDSFISQKSELIKM-NILSSLSRSVQVSNKEIDVAIL----- 156
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY     264 VDYDDLMLSL-----LKEAAKMKMINTCNEYOKR-----HGKKTLP 299
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB     157 SDQKDVETSMQVFTSKDGNKRAFTQNNLNKRNKKCADVKKSLYD 201
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 26
JC4298
hyaluronan receptor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 16-Nov-1995 #sequence revision 08-Feb-1996 #text_change 09-Jul-2004
C;Accession: JC4298; A41923; S21586
R;Entwistle, J.; Zhang, S.; Yang, B.; Wong, C.; Li, Q.; Hall, C.L.; A, J.; Mowat, M.; Gr
Gene 163, 233-238, 1995
one preceding the hyaluronan receptor RHAMM.

```

## RESULT 26

hyaluronan receptor - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 16-Nov-1995 #sequence revision 08-Feb-1996 #text\_change 09-Jul-2004  
 C;Accession: JC4298; A42925; A41923; S21586  
 R;Entwistle, J.; Zhang, S.; Yang, B.; Wong, C.; Li, Q.; Hall, C.L.; A, J.; Mowat, M.; Gr  
 Gene 163, 233-238, 1995  
 A;Title: Characterization of the murine gene encoding the hyaluronan receptor RHAMM.  
 A;Reference number: JC4298; MUID:96011639; PMID:7590272  
 A;Accession: JC4298  
 A;Molecule type: mRNA  
 A;Residues: 1-631 <ENT>  
 A;Cross-references: UNIPROT:Q00547; EMBL:X64550  
 A;Experimental source: 3T3 fibroblast  
 R;Hardwick, C.  
 J. Cell Biol. 118, 753, 1992  
 A;Reference number: A42925; MUID:92348516; PMID:1639856  
 A;Contents: erratum  
 A;Accession: A42925  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 155-376; 'S', 378-504, 'E', 506-631 <HAR>  
 A;Cross-references: GB:X64550  
 A;Note: authors translated the codon AGT for residue 377 as Thr and CTA for residue 507  
 R;Hardwick, C.; Hoare, K.; Owens, R.; Hohn, H.P.; Hook, M.; Moore, D.; Cripps, V.; Auste  
 J. Cell Biol. 117, 1343-1350, 1992  
 A;Title: Molecular cloning of a novel hyaluronan receptor that mediates tumor cell motil  
 A;Reference number: A41923; MUID:92299690; PMID:1376732  
 A;Accession: A41923  
 A;Status: preliminary  
 A;Molecule type: mRNA

A;Note: this sequence has been corrected in reference A42925

C; Genetics: 7.4%; Score 112; DB 2; Length 631;  
A; Gene: rhann 20.5%; Pred. No. 13;  
A; Introns: 20/3; 54/2; 79/2; 137/3; 187/3; 259/2; 382/2; 431/2; 482/2; 515/3; 574/3; 625/  
C; Superfamily: hyaluronan receptor 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;  
C; Keywords: glycoprotein; receptor  
F; 260-382/Region: 21 residue repeats  
F; 516-574/Region: hyaluronan binding #status predicted  
F; 575-625/Region: hyaluronan binding #status predicted  
F; 116,283,304,325,346,367,398,438,619/Binding site: carbohydrate (Asn) (covalent) #status  
Query Match  
Best Local Similarity  
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;  
1 MTEIVADKTVVKNALETADGALDLYNKLDQVTPWQTFDETIKLSRFKQRYQAASV 60  
QY

Db 253 LDNLLREKEVELEKHAHAQAIIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 309  
Qy 61 LVGDITKLLMDSQDKYFEATQTVYVCGVATQALAAAYILLDFBYNEKASQAQKILIKVL 120  
Db 310 L-RDVTAAQLESEQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLSLRDVTAQ-L 359  
Qy 121 DDGITKLENAQKSL-LVSSQSFNNASGKLAL-DSQLTN-DFSEKSYFOSQVDKIRKEA 177  
Db 360 ESVOEKYNDTAQSLRDVTAQLESYKSTLKEIEDLKLNTLQEKVAMAEKSVEDVQQOI 419  
Qy 178 YAGAAAGVAGPFLGISYAGAGVVEGKLIPELKNK-----LKSQVQFF---TTL 226  
Db 420 LTAESTNQ-----EYA-----RVQDLQNRSTLKEBEIKETSPLEKIDTLK 462  
Qy 227 NTVKQANKDI-----DAKLKLTTRIAAIGETTTETTRFRFYVDYDDL 269  
Db 463 NQLRQODEDFRQLEKRGKRTAKENVMTELT-----MEINKWRLLYDEL 507  
RESULT 27  
S38173  
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)  
N;Alternate names: protein YKR095w; protein YKR415  
C;Species: Saccharomyces cerevisiae  
C;Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 09-Jul-2004  
C;Accession: S38173; S40647; S31207  
R;Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo, J.  
submitted to the Protein Sequence Database, March 1994  
A;Reference number: S38158  
A;Accession: S38173  
A;Molecule type: DNA  
A;Residues: 1-1875 <BAL>  
A;Cross-references: UNIPROT:Q02455; EMBL:Z28320; NID:G486586; PID:G486587; MIPS:YKR095w  
A;Experimental source: strain S288C  
R;Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; J.  
Yeast 9, 1349-1354, 1993  
A;Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromo  
A;Reference number: S40644; MUID:94205265; PMID:8154186  
A;Accession: S40647  
A;Molecule type: DNA  
A;Residues: 1-1875 <BOU>  
A;Cross-references: EMBL:X73541; NID:G450550; PIDN:CAA51948.1; PID:G450554  
A;Experimental source: strain S288C  
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.  
Mol. Gen. Genet. 237, 359-369, 1993  
A;Title: A new yeast gene with a myosin-like heptad repeat structure.  
A;Reference number: S31207; MUID:93247549; PMID:8483450  
A;Accession: S31207  
A;Molecule type: DNA  
A;Residues: 1-300, 'A', 302-1875 <KOE>  
A;Cross-references: EMBL:L01992; NID:G171958; PIDN:AAA34783.1; PID:G171959  
A;Gene: SGD:MLP1  
C;Genetics:  
A;Cross-references: SGD:S0001803; MIPS:YKR095w  
A;Map position: 11R  
Query Match 7.4%; Score 112; DB 2; Length 1875;  
Best Local Similarity 19.7%; Pred. No. 52;  
Matches 68; Conservative 66; Mismatches 122; Indels 90; Gaps 15;  
Qy 2 TEIVADKTVVKNVNAIETAGALDLYNK---YLDQVI-----PWQTFDETIK----- 45  
Db 685 SRLAERFKLNTLTDLTKAENDQLKRPDYLQNTILKQDSKTHETINYSVCSKLSI 744  
Qy 46 ---ELSRFKOEY-----SQAAVLVDGDIKTLMDSDQKYFEATQ 82  
Db 745 VETELLNKEQKLRVHLEKNLQKLNKLSPEKDSLRIMWTQTLQTKEREDLLEETRKS 804  
Qy 83 VYVCGVATQALAAAYILLDFBYNE-----KASQAQKILIKVL-DDGITKLENAQKSLV 136  
Db 805 CQK-----KIDELEDALEKSKTSQKDHIIKQLEEDNNSNIWYQNKIEA 850  
Qy 137 SSQSFNASGKLALDSQLTN-----DFSEKSYFOSQVDKIRKEAYAGAAAGVAGPFLG 192

Db 851 LKDYISV---ITSVDSKQTDIEKLYKVKLSKEIEEDKIRLHTY-----NVM-----ET 899  
Qy 193 IISYSIAAGVVEGKL-IPELKNKLSQVNFPTLTSNTVKQANKDIDAAKLLKLTTEIAAIG 251  
Db 900 INDSLRLKELEKSKINLTDAYSQIKEYKDIYETTSQSLQQTNSKLDSEKDFDTQIKNLT 959  
Qy 252 EIKTETTRFRFYVDYDDLMLSLKLEAAKMMINTCNE--YKRRHGKK 295  
Db 960 DEKTSLED-----KISLKE---QMFNLNNELDLQKGMK 992  
RESULT 28  
S67087  
hypothetical protein YOR195w - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein O4806  
C;Species: Saccharomyces cerevisiae  
C;Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 09-Jul-2004  
C;Accession: S67087  
R;Hughes, B.; Pohl, T.M.  
submitted to the Protein Sequence Database, July 1996  
A;Reference number: S66685  
A;Accession: S67087  
A;Molecule type: DNA  
A;Residues: 1-821 <HUG>  
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:G1420464; PID:e252389; PID:G1420465  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: SGD:SLK19; MIPS:YOR195w  
A;Cross-references: SGD:S0005721  
A;Map position: 15R  
Query Match 7.4%; Score 111.5; DB 2; Length 821;  
Best Local Similarity 21.8%; Pred. No. 20;  
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;  
Qy 10 VEVVNAIETAGALDLYNKLYLDQVLPWQTFDETIKLSRPFQKYSQAASVL----- 61  
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430  
Qy 62 -----VGDIKTLMDSDQKYFEATQTVYVCGVATOLAAAYILLDFBYNEKASQAQKIL 116  
Db 431 DKVVELNMAELKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477  
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFOSQVDKIR 174  
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536  
Qy 175 KEAYAGAAAGVAGPFLGISYAGAGVVEGKLIPELKN-----KLKSVQNFPTLSTVK 230  
Db 537 K-----VWQDKLI---KNETLKLKEA-----ID 558  
Qy 231 QANKDIDAACKLTT--ETAAIGETITET-----ETTRFYVDYDDL----- 269  
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKESKLHEE 618  
Qy 270 MSLSLKEAAKMMINTCNEVYKRRHGKTLFEV 300  
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648  
RESULT 29  
T18296  
myosin heavy chain - Entamoeba histolytica  
C;Species: Entamoeba histolytica  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18296  
R;Guillen, N.  
submitted to the EMBL Data Library, February 1997  
A;Reference number: Z18865  
A;Accession: T18296  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA



A;Residues: 1-2139 <GUI>  
A;Cross-references: UNIPROT:Q07569; EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB480  
C;Genetics:  
A;Gene: mhca  
C;Superfamily: myosin heavy chain; myosin motor domain homology  
F;91-780/Domain: myosin motor domain homology <MMO>

Query Match 7.4%; Score 111.5; DB 2; Length 2139;  
Best Local Similarity 21.3%; Pred. No. 67; Mismatches 128; Indels 91; Gaps 15;  
Matches 74; Conservative 57;

QY 7 DKTEVVKNAIETAD-----GALDLYNKYLQVPIWQ-----TFDETIKELSRFKQY 54  
DB 986 DKTTAEQESIDEKEDITKLGDKILKEBKDDLEQDRADVATKDDIAKLNKITEC 1045  
QY 55 SQAASVLVGDITLLMDSQKIFATQTVIEW-----CGVATQLLAAYILLFDEYNEKASA 111  
DB 1046 EDARDE-IAKLEQLEDEENKNDLTNELOQTQLKGETEKSLAAQVAA-----TKASD 1099  
QY 112 QKDILIKVLD-----GITKLN-EAKSLVSSQSP-----NNASGKILA 150  
DB 1100 ERDLSQNLNEKUTTNLTITKADLEKKISGLKQDYEDLEDKXKIEGLDLNAQRKIKE 1159  
QY 151 LDSQLTNDFSEKSSYFQVDKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKLI-P 209  
DB 1160 LDDEITKG-ADVSQYLQKQ-----KEEYES-----QIAKNQBEKEAIGN 1197  
QY 210 ELKNKLKS-----VQNFFTLNTVQKANKDIDAAKILTTTIAIGIKETETTRF 262  
DB 1198 DVKNKEITKEKELEIQSLQKLDDET-EVEKEDAEKKKETEKEKMKALQEBENVESSKN 1256  
QY 263 YVDVDDLML-----SLKKEAAKKNMINTCNEYQKRHGK 294  
DB 1257 STEKDKKLEKLDKDTQKLDMDTADNEKAKAKALEAQLNEVDNHEK 1306

RESULT 30  
F87908  
Protein T22A3.8 [imported] - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
A;Accession: F87908; F87908  
R;anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A;Reference number: A75000; MUID:99069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A;Accession: F87908  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2823 <STO>  
A;Cross-references: UNIPROT:Q45614; GB:chr\_I; PIDN:CAA15432.1; PID:g3924779; GSPDB:GN000  
A;Accession: F87908  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2823 <ST2>  
A;Cross-references: GB:chr\_I; PIDN:CAB03385.1; PID:g3924881; GSPDB:GN00019; CESP:T22A3.8  
A;Gene: T22A3.8  
A;Map position: 1  
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.4%; Score 111.5; DB 2; Length 2823;  
Best Local Similarity 21.3%; Pred. No. 95;  
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

QY 25 DLVKNKYLDQVPIWQT-FDETIKELSRFKQYSAASVLVG-----DIKTLMD 72  
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADARMSLVVGRINRYKEVNSIEKURVEA 1939  
QY 73 QDK-----YFEATQTV-----YEWCGVATQLLAAYILLFDEY 104  
DB 1940 EDQIAYSRSNIEKARSELMNMFEDKEKINNTLAELPDLVEQCONITLL----YSQLIDEY 1996  
QY 105 NE-----KKASAQKIDILIKVLDGDKITKLENAQKSLVSS-----OSFNAS--- 145  
DB 1997 DEEYVQTAGRHAELKLEVQQAQ---KIVDRFVDTRTETENPLKASHAYENIVEALKNATEAV 2053  
QY 146 -----GKLALDSQLTNDPSEKSSYFQSDVKIRKEAYAGAAAGVAGVAGPFGLLIS 195  
DB 2054 DSAAEASEAVSKMLGSESGSDANEES--LRSOLEKLNES-----S 2094  
QY 196 YSIAAGVVEGKLIPELKNKLKSVQNFFTLNTVQKANKDIDAAKILTTTIAAIGIKT 255  
DB 2095 LSNVDSNAVKIVBELKKEKD-----LTDRGLHN-----ELK-TSIVKRLGVKN 2140

1940 EDQIAYSRSNIEKARSELMNMFEDKEKINNTLAELPDLVEQCONITLL----YSQLIDEY 1996  
105 NE-----KKASAQKIDILIKVLDGDKITKLENAQKSLVSS-----OSFNAS--- 145  
1997 DEEYVQTAGRHAELKLEVQQAQ---KIVDRFVDTRTETENPLKASHAYENIVEALKNATEAV 2053  
146 -----GKLALDSQLTNDPSEKSSYFQSDVKIRKEAYAGAAAGVAGVAGPFGLLIS 195  
2054 DSAAEASEAVSKMLGSESGSDANEES--LRSOLEKLNES-----S 2094  
196 YSIAAGVVEGKLIPELKNKLKSVQNFFTLNTVQKANKDIDAAKILTTTIAAIGIKT 255  
2095 LSNVDSNAVKIVBELKKEKD-----LTDRGLHN-----ELK-TSIVKRLGVKN 2140  
256 ETETRFFVYDDMLMLLLKEAAKKM-INTCNEYOKRHGKKT 296  
2141 EASS---WDDKDHMSILKNGAKTAHARSANVKESEGIKT 2179

RESULT 31  
T23064  
hypothetical protein T22A3.8 - Caenorhabditis elegans (fragment)  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T23064; T25096  
R;Barlow, K.  
submitted to the EMBL Data Library, October 1997  
A;Reference number: Z19669  
A;Accession: T23064  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2823 <WIL>  
A;Cross-references: UNIPROT:Q45614; EMBL:AL008585; PIDN:CAA15432.1; GSPDB:GN00019; CESP  
A;Experimental source: clone H10E24  
R;McMurray, A.  
submitted to the EMBL Data Library, October 1996  
A;Reference number: Z19980  
A;Accession: T25096  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2823 <W12>  
A;Cross-references: EMBL:Z81125; PIDN:CAB03385.1; GSPDB:GN00019; CESP:T22A3.8  
A;Experimental source: clone T22A3  
C;Genetics:  
A;Gene: CESP:T22A3.8  
A;Map position: 1  
A;Introns: 45/1; 282/2; 312/3; 416/2; 1255/3; 1329/3; 1418/3; 1776/2; 1989/2; 2760/2  
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

Query Match 7.4%; Score 111.5; DB 2; Length 2823;  
Best Local Similarity 21.3%; Pred. No. 95;  
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

QY 25 DLVKNKYLDQVPIWQT-FDETIKELSRFKQYSAASVLVG-----DIKTLMD 72  
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADARMSLVVGRINRYKEVNSIEKURVEA 1939  
QY 73 QDK-----YFEATQTV-----YEWCGVATQLLAAYILLFDEY 104  
DB 1940 EDQIAYSRSNIEKARSELMNMFEDKEKINNTLAELPDLVEQCONITLL----YSQLIDEY 1996  
QY 105 NE-----KKASAQKIDILIKVLDGDKITKLENAQKSLVSS-----OSFNAS--- 145  
DB 1997 DEEYVQTAGRHAELKLEVQQAQ---KIVDRFVDTRTETENPLKASHAYENIVEALKNATEAV 2053  
QY 146 -----GKLALDSQLTNDPSEKSSYFQSDVKIRKEAYAGAAAGVAGVAGPFGLLIS 195  
DB 2054 DSAAEASEAVSKMLGSESGSDANEES--LRSOLEKLNES-----S 2094  
QY 196 YSIAAGVVEGKLIPELKNKLKSVQNFFTLNTVQKANKDIDAAKILTTTIAAIGIKT 255  
DB 2095 LSNVDSNAVKIVBELKKEKD-----LTDRGLHN-----ELK-TSIVKRLGVKN 2140



QY 256 ETETTRFYVDYDMLSLKEAAKM-INTCNEYQKRGKKT 296  
Db 2141 EASS---WDDKDRMHSILKNGAKTAHSAHVKKSEGIKT 2179

RESULT 32  
T43291  
laminin alpha chain - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T43291  
R;Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang  
submitted to the EMBL Data Library, June 1998  
A;Description: Expression, function and evolution of laminin alpha chains.  
A;Reference number: Z22397  
A;Accession: T43291  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-3102 <ZHU>  
A;Cross-references: UNIPROT:O45614; EMBL:AF074902; PIDN:AAC26793.1  
C;Genetics:  
A;Map position: 1  
A;Note: lamal/2  
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.4%; Score 111.5; DB 2; Length 3102;  
Best Local Similarity 21.3%; Pred. No. 1.1e+02;  
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

QY 25 DLVNYKLYDVPWQT-FDETIKELSRFKQESQAASVLVG-----DIKTLMDS 72  
Db 1881 DLKNR-IDVLEQWMDYRTIYDVKKTDADAEKMSLVVGRINRYKEVSNIEKLRVEA 1939

QY 73 QDK-----YFPAQTQV-----YEWCGVATQLLAAYILLFDEY 104  
Db 1940 EDQIAYSRNSTEKARSELMNMFEDKEKNWTLAELPDLVEQCQNITLL---YSQLIDEX 1996

QY 105 NE-----KASAKDILIKVLDDGGITKLENAOKSLVSS-----QSFNAG--- 145  
Db 1997 DEEYVQTAGRHAELVQAO---KIVDRFVDTRTETENPLKASHAYENIVEALKNATEAV 2053

QY 146 -----GKLLALDSQLTNDPSEKSSYFQSOVDKIRKEAYAGAAAGVAGPFGLIIS 195  
Db 2054 DSAEASEAVSKMGSGESGDANEE---LRSOLEKLNES-----S 2094

QY 196 YSTAAGVVEGKLIPELKNKLKSVQNFPTTSLNTVKQANKDIDAAKLKLTTEIAAIGIKT 255  
Db 2095 LSNVDNSNAKIVPELKKKKO-----LTDRLGLN-----ELK-TSIVKELGVKN 2140

QY 256 ETETTRFYVDYDMLSLKEAAKM-INTCNEYQKRGKKT 296  
Db 2141 EASS---WDDKDRMHSILKNGAKTAHSAHVKKSEGIKT 2179

RESULT 33  
H88391  
protein R06B10.2 [imported] - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C;Accession: H88391  
R;anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A;Reference number: A75000; MUID:99069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C.ele  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A;Accession: H88391  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-934 <STO>  
A;Cross-references: UNIPROT:O44874; GB:chr\_III; PIDN:AAB95035.1; PID:g2746881; GSPDB:GNC  
C;Genetics:  
A;Gene: R06B10.2

A;Map position: 3

Query Match 7.4%; Score 111; DB 2; Length 934;  
Best Local Similarity 21.6%; Pred. No. 25;  
Matches 69; Conservative 43; Mismatches 131; Indels 76; Gaps 14;

QY 5 VADTVEVVKNAIETADGALDLYNKYLDQV-IPQWTFDETIKELSRFKQESQAASVLVG 63  
Db 81 VYEKTGVPGSIPKPD---DVYKTQDALKIKWSKVDSPAKGIVIFLNEMESMEGVTS 137

QY 64 DIKTLMDSDQKFEATQTVVYECGATQLLAAYILLFDEYNEKKASAKDILIKVLDDG 123  
Db 138 DDANIKLSDMADPHLPFKF---DW-----TLEEN----- 163

QY 124 ITKLENAOKSLVSSQSFN---NASGKLL-----ALDSQLTNDPSEKSSYFQSOVDKIRKE 176  
Db 164 VYAQNQHLRNIHISYPSLNLVYKAVAFPPYKNSLDGSLLPQVENINKLNTLAEKIKPE 223

QY 177 AYAGAAAGVAGPFGLIISYISIAAGVVEGKLIPELKNKLKSVQNFPTTSLNTVKQANKDI 236  
Db 224 -----APMISQLHELFLSRLKKS-VNFKIISGLLNGISDIQIMFTDLKN--EWVNKT 274

QY 237 DAAKLKLTTEIAAIGETKTETTRFYVDYDMLSLKEAAKKNMINTCNEYQKRGKKT 296  
Db 275 DGQAYNLG---KALNQKLTGLNNAQ-QVD-----LSIVPYEEKIFKDVSNLYQKVNLSKD 325

QY 297 L-----FEVPE 302  
Db 326 LTVIVRPLIGTVTGIBIPQ 344

RESULT 34  
LEECA  
hemolysin A - Escherichia coli  
C;Species: Escherichia coli  
C;Date: 30-Sep-1988 #sequence\_revision 30-Sep-1988 #text\_change 09-Jul-2004  
C;Accession: A24433; I41280  
R;Feilmee, T.; Pellett, S.; Welch, R.A.  
J. Bacteriol. 163, 94-105, 1985  
A;Title: Nucleotide sequence of an Escherichia coli chromosomal hemolysin.  
A;Reference number: A24433; MUID:85234404; PMID:3891743  
A;Accession: A24433  
A;Molecule type: DNA  
A;Residues: 1-1023 <PEL>  
A;Cross-references: UNIPROT:P09983; GB:M10133; GB:M12863; NID:g146377; PIDN:AAA23975.1;  
A;Experimental source: strain J96, O4 serotype  
R;Stanley, P.; Packman, L.C.; Koronakis, V.; Hughes, C.  
Science 266, 1992-1996, 1994  
A;Title: Fatty acylation of two internal lysine residues required for the toxic activity  
A;Reference number: A55387; MUID:95099325; PMID:7801126  
A;Contents: annotation; lysine palmitoylation  
A;Note: lysine modification is performed by the hlyC gene product  
R;Haertlein, M.; Schiesl, S.; Wagner, W.; Rdest, U.; Krest, J.; Goebel, W.  
J. Cell Biol. 22, 87-97, 1983  
A;Title: Transport of hemolysin by Escherichia coli.  
A;Reference number: I41280  
A;Accession: I41280  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1,'T',3,'V',5,'T',7-44 <RES>  
A;Cross-references: GB:M29173; NID:g146337; PIDN:AAA23957.1; PID:g146338  
C;Genetics:  
A;Gene: hlyA  
C;Function:  
A;Description: attacks blood cell membranes and causes cell lysis  
C;Superfamily: hemolysin A; hemolysin A homology  
C;Keywords: calcium binding; cytolysis; exotoxin; hemolysis; lipoprotein; tandem repeat;  
F;246-791/Domain: hemolysin A homology <HLYA>  
F;723-851/Region: 9-residue repeats (G-G-X-G-[DN]-D-X-[LVIVYF]-X)  
F;563,689/Binding site: palmitate (Lys) (covalent) #status experimental

Query Match 7.4%; Score 111; DB 1; Length 1023;  
Best Local Similarity 21.3%; Pred. No. 28;

```

193 DIVLP--AYDSSKLLSFNELRPERLVAQAATRONNEBEGEKMNKPKVPFELSHSDVRTL 250
Db

70 -----MDSQKYFEATQTVYEWCGV-----ATQLLAAYILLFDE 103
QY

251 YSLDPTLDKESDFTADVDTYF--IVNVYWSMDENSYDKYKFNYYRTGDTVSIDFDS 308
Db

104 YNE-----KKASAQDILIKVLDDGITKLNEAQKSLLYSSQSFNNASGKL 148
QY

309 KTEVMTNWEVVPPIQSLNKKQBEQIKDLTKQVNLKDKVGI---EQQNTASEKL 365
Db

149 LALDSQ-----LNTDFSEKSYFQSOVDKIRKEAVAGAAAGVVGPPGLI 193
QY

366 VQLNSEVEQLPKYKEKHEKTLLEQKLSEKNFEYKAFPEALNAEE----- 409
Db

194 ISYSTAGVVEGKLIPELNKLSVQNFFTLSTNTVKQANKDIDAAKLKLTEIAAIGI 253
QY

410 -----KFTSEVQNL-----IHASVKQ--DEGEKAVQLQNTMLVDLYSV 447
Db

254 KTETETT 260
QY

448 PTEIWT 454
Db

RESULT 36
T44967
gas-vesicle protein gvpC - Halorubrum vacuolatum

```

C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #seq\_change 0  
 C;Accession: T44967  
 R;Mayr, A.; Pfeifer, F.  
 Arch. Microbiol. 168, 24-32, 1997  
 A;Title: The characterization of the gypACNOFGH gene cluster involved in gas vesicle formation in *Halobacterium salinarum* R1  
 A;Reference number: Z22885; MUID:97361906; PMID:9211710  
 A;Accession: T44967  
 A;Status: preliminary; translated from GE/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-540 <MAY>  
 A;Cross-references: UNIPROT:Q33398; EMBL:Y08571; NID:q2598010; PIDN:CAA69882.1  
 A;Experimental source: strain DSM 3638  
 C;Genetics:  
 C;Gene: gypC

|    |     |                               |                                     |                          |            |                                    |
|----|-----|-------------------------------|-------------------------------------|--------------------------|------------|------------------------------------|
|    |     | Query Match                   | 7.3%;                               | Score 110.5;             | DB 2;      | Length 540;                        |
|    |     | Best Local Similarity         | 20.9%;                              | Pred. No. 13;            |            |                                    |
|    |     | Matches                       | 71;                                 | Conservative             | 42;        | Mismatches 130; Indels 97; Gaps 12 |
| Qy | 5   | VADKTVEV---                   | KNAIETADGALDLYNKYLDOVI              | PWGTDETIKELSRFQEVSQAASVL | 61         |                                    |
|    |     | :::::                         | :::::                               | :::::                    |            |                                    |
| Dd | 3   | VNEKEEMVAKSEFVEVQRAFTVAAS     | FADDVEQODTSSQVASIAAPREEMQQSSAVF     | 62                       |            |                                    |
|    |     | :::::                         | :::::                               | :::::                    |            |                                    |
| Qy | 62  | V-----                        | GRIKTL---                           | LMSDDKYFEATQTIVYEW-      | CGVATOLLAA | 96                                 |
|    |     | :::::                         | :::::                               | :::::                    |            |                                    |
| Dd | 63  | ESYSEBFDGVNRLNAIEAQODAFEA     | TAVAFEYRESFHOOVQALLTAIGELQQOFVE     | 122                      |            |                                    |
|    |     | :::::                         | :::::                               | :::::                    |            |                                    |
| Qy | 97  | YILLFDEYNKKASAOKITILIKVLDDG   | DTKLNEAQSLLVSSQSFNNASGKLIALDSQL-    | 155                      |            |                                    |
|    |     | :::::                         | :::::                               | :::::                    |            |                                    |
| Dd | 123 | VEDVFQYAGDFGEAEVDENAEVSDRIA   | FENTAAFEYDKTHREKVSGALLAAVDLR        | 182                      |            |                                    |
|    |     | :::::                         | :::::                               | :::::                    |            |                                    |
| Qy | 156 | -----                         | TNDFSEKSYFOQOVDVKIRKEAYAGAAGVGAGPP- | 190                      |            |                                    |
|    |     | :::::                         | :::::                               | :::::                    |            |                                    |
| Dd | 183 | ARFANVTAAFDQYAEDFQGSVDFFNS    | KVDDQR-----                         | APEITAAEFSEYRDEFHQDV     | 236        |                                    |
|    |     | :::::                         | :::::                               | :::::                    |            |                                    |
| Qy | 191 | -GLIIYSISIAAGVVGEKLIPELKNKLKS | VQNPFLLTSNTVVOANKDID--              | AAKLKLITE-               | 246        |                                    |
|    |     | :::::                         | :::::                               | :::::                    |            |                                    |
| Dd | 217 | PGHJLA-----                   | IIELEORPSVDABEFGRYAD----            | EFGVEVDDDAASFEDTTAA      | 281        |                                    |
|    |     | :::::                         | :::::                               | :::::                    |            |                                    |

```

247 -----IAATGEIKTETTR--FVVDYDD 268
      ||| | : | |
282 FEEYRDKFRQEQVAVLAIDLENDIENATRAEFATD 321

```

bps2 protein homolog (bps2) [imported] - Sulfolobus solfataricus  
C;Species: Sulfolobus solfataricus  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C;Accession: A90394  
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A;Description: Sulfolobus solfataricus complete genome.  
A;Reference number: A99139  
A;Accession: A90394  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-587 <KUR>  
A;Cross-references: UNIPROT:Q97WH8; GB:AE006641; NID:gl3815540; PIDN:AAK42408.1; GSPDB:G  
C;Genetics:  
A;Gene: bps2

Query Match 7.3%; Score 110; DB 2; Length 587;  
Best Local Similarity 20.6%; Pred. No. 16;  
Matches 59; Conservative 56; Mismatches 106; Indels 66; Gaps 12;

Qy 23 ALDLYNKYLQVTPWQTFDETI--KELSRFKQYSQAASVLVGDITKLLMDSQDKYPEAT 80  
Db 51 AEDLLNVFADSGVYEALDNKLYRRIKRIIRNGLGEKKNLMDDRRALL--TYSPE 106  
Qy 81 QTVYEWCGVATQALAA-----YILLFDEYNEKASAKOILIKVLDDGTITKLENAQKSL 134  
Db 107 NR-----LVTQILSGDGVWFSTTSKINEIK--AKBELQKLLTAENARDELQK-- 156  
Qy 135 LVSSQSENN--ASGLLADSQLTNDFSEKSSVFQSDVKIRKEAYAGAAAGVAGPFG 191  
Db 157 -----KYNRIEIOAKIRAD-----BEIDKLEKB--RESSNIVAK-- 191  
Qy 192 LIISYSIAAGVVEGKIPKLNKLSVQNPFTLSTNTVKQAKDIDAAKLKTETIAAIG 251  
Db 192 --TYYTIT--LTRQNKINEILNKIKVKDELANLEFALKIEIEIQNESKVSFDIKTL 247  
Qy 252 EIKTETETTFRYVDYD-----LMLSLKKAQKMWNTCN 286  
Db 248 EKEMEENEXKLKTNDRSELETBLKVLRLVLEEVNESDRHLDTCN 294

RESULT 38  
G97236  
ATPase involved in DNA repair [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: G97236  
R;Molling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: G97236  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1163 <KUR>  
A;Cross-references: UNIPROT:Q97FKI; GB:AE001437; PIDN:AAK80682.1; PID:gl5025772; GSPDB:G  
C;Genetics:  
A;Gene: CAC2736

Query Match 7.3%; Score 110; DB 2; Length 1163;  
Best Local Similarity 17.9%; Pred. No. 38;  
Matches 73; Conservative 69; Mismatches 114; Indels 152; Gaps 14;

Qy 1 MTEIVADKTVEVVKNAIETADG-----ALDLYNKYLQVTPWQTFDETIK 45  
Db 109 IVDITGDE-VEVLEBEGAKSVNEKQIEIIGLSLDFTRTV--VLPQGFSEFLKLEGGKERR 165  
Qy 46 -----ELSRFKQYSQAASVLVGDITKLLMDSQ 73

Query Match 7.3%; Score 110; DB 2; Length 3187;  
Best Local Similarity 20.8%; Pred. No. 1.4e+02;  
Matches 80; Conservative 65; Mismatches 114; Indels 126; Gaps 18;

Qy 1 MTEIVADKTVE--VVKNA-IETADGALDLYNKYLQVTPWQTFDETIKLSRFKQYSQ 56  
Db 1725 VTEAVVKSQESQSLSENKLEDAEATL-LANSAKPGVS--ETP-SSHDDINNYLQQLDQ 1780  
Qy 57 AASVLVGDITKLLMDSQDKYFEATQTVYEWCGVATQALAAAYILLFDEYNEKKA-----SAQ 112  
Db 1781 ----LKGRIAELEMEKQ-KDRELSQTL-----ENEKNALLTQISA 1815  
Qy 113 KDILIKVLDDGITKLN-----EAKSLVSSQSFNNASGKLLALDSQTNDFSEKS-----S 164  
Db 1816 KDSLEKLLBEVAKINMLNQIQEISLVRVTKLETABEEKDDLEERLMNLQAEIENLGN 1875  
Qy 165 YFQSQVD-----KIRKAYAGAAAGV 186  
Db 1876 YYQDVDTAQIKNQLSESEMQLKRCVSELEEBEKQQLVKEKTKVSEIRKE-YMEKIQAG 1934  
Qy 187 AGPFGILISYSIAAGVVEGKLIPELKNKLSVQ-----NFFTTLSNTVK-----QA 232  
Db 1935 KGFGSKIHAKELQ-----ELLKKEQEVKQLQKDCIRYLGRISALEKTKVAKLFBVHTES 1988  
Qy 233 NKDIDAAK-----LKTETIAAIGIKETETETTRFYVDYDML----- 271  
Db 1989 QKLDATKGNLAQAVEHHKKAQAEISFKILLDDTQSEAAARVLADNLKJLKEQLQNKESI 2048  
Qy 272 -SLLKEAAKMWNTCNEYQKRGHK 295

Db 166 NMLERLNLQYGDLSFLKARKIRKERKENVLGKGVENINEDVLKERRELLKNN 225  
Qy 74 DKYFEATQTVYEWCGVATQALAAAYILLFDRYN-----EKASAKQDILIKV-- 119  
Db 226 DPFNEASK-----BYLKAEBEYNEGKEVWGLQIEIEEKRNVRKOLMEKDE 271  
Qy 120 --LDDGTTKLENAQKSLVSSQSFNNASGKLLALDSQTN-----DFSEKSS 164  
Db 272 IDLKEKRLGESSKVKPYIDNYENTLKGIDILKEQILSRENTMKAISLEKEDMEKLS 331  
Qy 165 YFQSQVDKIRKEAYAGAAAGVAGVVEGKLIISYSIAAGVVEGKLIPELKNKLSQNFPTT 224  
Db 332 IAKONKEK-----ALPKFMIKHIIILDAIKEKDLIDNLEKRLQGGKIEK 377  
Qy 225 LS-----NTVQAKNDIDAQKLLKTTETAAIGETETE----- 258  
Db 378 LSLEASNKEELIKQNTKIDSLTLKIQNLSEKIDNLKVPPEYKKNKINEGIFLLRNYDEKL 437  
Qy 259 -----TTRFVVDYD-----DLMLSLKKAQKMWNTCNEYQK 291  
Db 438 KHKNKLGDCRFQVDFEKAQSKKEMLFNKLEBERSKL-----DTYTKK 481

RESULT 39  
JC5837  
364K Golgi complex-associated protein - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 05-Mar-1998 #sequence\_revision 13-Mar-1998 #text\_change 09-Jul-2004  
C;Accession: JC5837  
R;Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.  
Cell Struct. Funct. 22, 565-577, 1997  
A;Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec  
F;49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict  
F;3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 7.3%; Score 110; DB 2; Length 3187;  
Best Local Similarity 20.8%; Pred. No. 1.4e+02;  
Matches 80; Conservative 65; Mismatches 114; Indels 126; Gaps 18;

Qy 1 MTEIVADKTVE--VVKNA-IETADGALDLYNKYLQVTPWQTFDETIKLSRFKQYSQ 56  
Db 1725 VTEAVVKSQESQSLSENKLEDAEATL-LANSAKPGVS--ETP-SSHDDINNYLQQLDQ 1780  
Qy 57 AASVLVGDITKLLMDSQDKYFEATQTVYEWCGVATQALAAAYILLFDEYNEKKA-----SAQ 112  
Db 1781 ----LKGRIAELEMEKQ-KDRELSQTL-----ENEKNALLTQISA 1815  
Qy 113 KDILIKVLDDGITKLN-----EAKSLVSSQSFNNASGKLLALDSQTNDFSEKS-----S 164  
Db 1816 KDSLEKLLBEVAKINMLNQIQEISLVRVTKLETABEEKDDLEERLMNLQAEIENLGN 1875  
Qy 165 YFQSQVD-----KIRKAYAGAAAGV 186  
Db 1876 YYQDVDTAQIKNQLSESEMQLKRCVSELEEBEKQQLVKEKTKVSEIRKE-YMEKIQAG 1934  
Qy 187 AGPFGILISYSIAAGVVEGKLIPELKNKLSVQ-----NFFTTLSNTVK-----QA 232  
Db 1935 KGFGSKIHAKELQ-----ELLKKEQEVKQLQKDCIRYLGRISALEKTKVAKLFBVHTES 1988  
Qy 233 NKDIDAAK-----LKTETIAAIGIKETETETTRFYVDYDML----- 271  
Db 1989 QKLDATKGNLAQAVEHHKKAQAEISFKILLDDTQSEAAARVLADNLKJLKEQLQNKESI 2048  
Qy 272 -SLLKEAAKMWNTCNEYQKRGHK 295

Db 2049 KSOIKQKDEDLRLRLEQAEKHKE 2073

RESULT 40

Minor capsid protein 1608 - Lactobacillus phage phi-gle  
C;Species: Lactobacillus phage phi-gle  
C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 13-Aug-1999  
C;Accession: T13216  
R;Kodaira, K.I.; Oki, M.; Kakikawa, M.; Watanabe, N.; Hirakawa, M.; Yamada, K.; Taketo, Gene 187, 45-53, 1997  
A;Title: Genome structure of the Lactobacillus temperate phage phi gle: the whole genome  
A;Reference number: Z17631; MUID: 97225795; PMID: 9073065  
A;Accession: T13216  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1608 <KOD>  
A;Cross-references: EMBL:X98106; NID: e917136; PID: e247162; PIDN: CAA66745.1  
C;Genetics:  
A;Note: Rozf1608

Query Match 7.3%; Score 109.5; DB 2; Length 1608;  
Best Local Similarity 19.6%; Pred. No. 62;  
Matches 76; Conservative 67; Mismatches 137; Indels 129; Gaps 16;

Qy 1 MTEIVADKTVVEVKNNAIETADGALDLYNKVDQVWPQTFTDTIKELSRFKQEYSQAASV 60  
Db 412 LVELI--KYVNAHENTIVDIIGNLKGILGIKTV-WKTFSDIVYDIAMK-----FG 460

Qy 61 LVGDIKTLMDSDQKYFEATQTVVEMCGVATQLLAAVILLFD-----EYNEKKAQAQDI 115  
Db 461 LVGEKAQESDPLDKIDDALNLSKNOELIENLTKAFIAMPALKKGMEFIGMLASLRKSL 520

Qy 116 L-----TKVLD-----DGTITLNEAQAQLLVSSQFNNAS-----GKLLALDSQLTND 158  
Db 521 IETAAVSKMVDLFGSGVTSAGGKAVTQTVAKEAGGTAATAGSSKVLGRFPAKGGAISTA 580

Qy 159 FSEKSS-----YFQSQVDKIRKE-----AYA 179  
Db 581 ELEAASGLGGKAMMAARGLTKAVPYMSIAASIPELFGTTQKTLGKLGFPAGSAGGPAA 640

Qy 180 GAAA-----GVVAGPGLIISYIAAGVVEG--KLPELKNKLSV-----218  
Db 641 GAAAGSAMPVYVAVGVVIGLAGSLQGVSGSIQKGIKSPFKLTSKMSDLGHDMAK 700

Qy 219 -----QNFPTLSNTVKQANKDIDAALKLTTEIAAIGEIKTETTT-----260  
Db 701 KFSGSPKPSLNDKQFSKVSYSITKLNKO---AKIKITDTSGISKAQKLTDTTYCKM 757

Qy 261 -----RFY-----VDYDML--SLLKBAKMKMINTCNEYQKRHK 294  
Db 758 KKSVDKYGHKRWMSIKDYATLVQNGSMTEKEANKLNKAKENYNKQAK 806

RESULT 41

A64465  
Hypothetical protein MJ1322 - Methanococcus jannaschii  
C;Species: Methanococcus jannaschii  
C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C;Accession: A64465  
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.J.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; rsion, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996  
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C. A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A;Reference number: A64300; MUID: 96337999; PMID: 8688087  
A;Accession: A64465  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1005 <BUL>  
A;Cross-references: UNIPROT:Q58718; GB:U67572; GB:L77117; NID:g1591958; PIDN:AAB99331.1;  
C;Genetics:

A;Map position: REV1273394-1270377  
C;Superfamily: hypothetical protein MJ1322

Query Match 7.2%; Score 109; DB 2; Length 1005;  
Best Local Similarity 19.1%; Pred. No. 37;  
Matches 58; Conservative 55; Mismatches 93; Indels 98; Gaps 11;

Qy 7 DKTVEVVKNAIETADGALDLYNKVDQVWPQTFTDTIKELSRF-----TKELSRF-----KQEYSQ 56  
Db 605 DEILEDIKSQLNKFK--NPNQYLSAVSYLSNVDDEGIRNRIKEIENIVSGWKNKCKRE 661

Qy 57 AASVL-----VCDIKTLMDSDQKYFEATQTVVEMCGVATQLLAAVILLFDVYNEKKA 110  
Db 662 ELNKLREDEIREINRLKDLNKLNLNKEKE-----LIEIENRRSLKFDKYKEYLGL 710

Qy 111 AQDKILIKVLDDGITKLINEAQAQLLVSSQFNNASGKLLALD-----152  
Db 711 TEKLEELKNIKDGLSEI-----YINCSKILLAIIDNIRKYNKEDIEIYLNK 757

Qy 153 -----SQTNDSEKSSYFQSQVDKIRKEAYAGAAAGVAGPGLIISYSTAAGVVEGKLI 208  
Db 758 ILEVNKEINDIEERISVYNOKLDE-----INYN-----EEHKKI 792

Qy 209 PEL-----KXKLKXQVQFFTLSTVVKQANKDIDAALKLTTEIAAIGEIKTETTRFYV 264  
Db 793 KELYENKQELDNVRBQKTEIETGIETLYKQVESLKARL-----KEMSNLEKEKEKLTQFV 848

Qy 265 DYDD 268  
Db 849 EYLD 852

RESULT 42

S54091  
Hypothetical protein YPR070w - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein YP9499.25  
C;Species: Saccharomyces cerevisiae  
C;Date: 08-Jul-1995 #sequence\_revision 19-Oct-1995 #text\_change 09-Jul-2004  
C;Accession: S54091; S69058  
R;Badcock, K.; Churcher, C.M.  
Submitted to the EMBL Data Library, May 1995  
A;Reference number: S54059  
A;Accession: S54091  
A;Molecule type: DNA  
A;Residues: 1-566 <BAD>  
A;Cross-references: UNIPROT:Q12321; EMBL:Z49219; NID:g805025; PID:g805050; MIPS:YPR070w  
A;Experimental source: strain AB972  
R;Couch, J.  
Submitted to the EMBL Data Library, March 1996  
A;Description: The sequence of S. cerevisiae cosmid 9513.  
A;Reference number: S69057  
A;Accession: S69058  
A;Molecule type: DNA  
A;Residues: 1-566 <COU>  
A;Cross-references: EMBL:U51033; NID:g1230676; PID:g1230678; MIPS:YPR070w  
C;Genetics:  
A;Gene: SGD:MED1  
A;Cross-references: SGD:S0006274  
A;Map position: 16R

Query Match 7.2%; Score 108.5; DB 2; Length 566;  
Best Local Similarity 23.9%; Pred. No. 19;  
Matches 78; Conservative 43; Mismatches 131; Indels 75; Gaps 18;

Qy 17 IETADGALDLYNKY-----LDQVWPQTFTDTIKELSRFKQEYSQAASVLV- 62  
Db 8 VETLDSMIELFKDYKPGSITLENITRLCQTLGLESFTELSNELSL-----STASKIIV 63

Qy 63 -----GDIKTLMDSDQK--YFEATQTVVEMCGVATQLLAAVILLFDVYNEKKA 109  
Db 64 DVDYNNKQDRIDQVKLVLASNFQNFQDGEHEKSNILLNLSLTKYPDLKAFHNLKF 123

Qy 110 SAQDKILIKVLDDGITKLN-EAQSLLVSSQFNNASGKLLALDSQLTNDPSEKSSY---165

Db 124 LYLDAYSHIESDSTSHNGSSDKSLDSSNASFNN-OGKL-----DLFKYFTELUSHVIRQ 177  
Qy 166 -FOSQV--DKIRKEAYAGAAAGVAGPGLIISYIAAGVVEGKLIPELK-----NKLKS 217  
Db 178 CFQDNCCDFKVRTN-----LNDKFGI---YILTOG-INGKEVPLAKIYLEENKSDS 224  
Qy 218 VQNFHTL--SNTVKQANKDID--AAKLKLTETTAAGTAEIKTETTRFYVDY----- 266  
Db 225 QYRFYIYQETKSTWNEAENFNGISLWMEIVANAENSNYTDLIWFPEDFISPELII 284  
Qy 267 DDLMLSLKKAAMKMIN--TCNEYOKR 291  
Db 285 DKVTCSSNSSSPPIIDLFNNVNSR 311

RESULT 43  
H69378  
conserved hypothetical protein AF1032 - Archaeoglobus fulgidus  
C;Species: Archaeoglobus fulgidus  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C;Accession: H69378  
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
; Fleckmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae  
A;Reference number: A69250; MUID:98049343; PMID:9389475  
A;Accession: H69378  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-886 <KLE>  
A;Cross-references: UNIPROT:O29230; GB:AE001032; GB:AE000782; NID:Q2689355; PIDN:AA89021  
C;Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 7.2%; Score 108.5; DB 2; Length 886;  
Best Local Similarity 21.0%; Pred. No. 33;  
Matches 74; Conservative 59; Mismatches 109; Indels 111; Gaps 15;

Qy 8 KTVVVKNAIETADGALDLYNKYLDQVVPQTFDETIKELSRPKQBYSQASVLVGIKT 67  
Db 425 KAVEELKSARTC-----PVCGRLEDEHRKNIMAYETREMKRIAELAKA-----DEIK 475  
Qy 68 LL---MDSQDKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKASAKQDILIKVLD--D 122  
Db 476 KLERLEKVEKALEKQETVLKYRQVMDLXALENEL--SSHDAEKLKSAESSEYRKVKERLD 534  
Qy 123 GITKLENAQKSLIVSSQSFNNASGKLALDLSQTLNDFSEKSSYFQSQVDKIRKEAYAGAA 182  
Db 535 G---LRQOKILLSSASRIKELKSSLSREIIEALKNVESERGELHR-----KIREEGF--- 583  
Qy 183 AGVVAGPGLIISVIAAGVVEGKLIPELAKNKLKSVQNF-----TTLSNTVKOA 232  
Db 584 -----ESLELEREVQSLRPFFYNKWLKADSSRLESELEKRR 620  
Qy 233 NK---DIDAAKLKL-----TTEIAAIGEI 253  
Db 621 EKLEDEISEIAKLEANGAEIRGQIDELLRIYSEEHRRSLDEHLRKSKELAG---L 677  
Qy 254 KTTETTRFYVDYDDLMLSLKKAAMKMINTCNEYOKRHKGTLPF---VPEV 303  
Db 678 KSRLETRESLQSAEKDLKLEELQAKW---DEYRK---KVEVFEXIAIPEL 723

RESULT 44  
T44802  
type I site-specific deoxyribonuclease (EC 3.1.21.3) chain hsdR [imported] - Streptococ  
N;Alternate names: type IC restriction subunit  
C;Species: Streptococcus thermophilus  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T44802

R;Solow, B.T.; Somkuti, G.A.  
submitted to the EMBL Data Library, August 1999  
A;Description: Comparison of low molecular weight heat stress proteins encoded on plasmid  
A;Reference number: Z22847  
A;Accession: T44802  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1025 <SOL>  
A;Cross-references: UNIPROT:Q9RNW2; EMBL:AF177167; PIDN:AAF04356.1  
A;Experimental source: strain ST135  
C;Genetics:  
A;Gene: hsdR  
A;Genome: plasmid pER35  
C;Superfamily: type I site-specific deoxyribonuclease chain hsdR  
C;Keywords: hydrolase

Query Match 7.2%; Score 108.5; DB 2; Length 1025;  
Best Local Similarity 20.1%; Pred. No. 40;  
Matches 80; Conservative 52; Mismatches 109; Indels 157; Gaps 19;

Qy 7 DKTVEVVKNAIETADGALDLYNKYLDQVVPQTFDETI-----KELSRPKQBYSQAS- 59  
Db 628 DELVEIMKE-----YEKQFD-ASPYQ--DEKLYNQNKRLARKEKQYQSDGQW 673  
Qy 60 ---VLVCD-----IKTLMDSDQKY-----FEATQTVYEWCGVATQLLAA- 97  
Db 674 LDFVIVVDRLLTGDSPTQTLIDREMNYOKLQAFSRTNRIY--TGKDSGLIVSRKP 731  
Qy 98 -----ILLFDEYNE-KKASAKQDILIKVLDGDTIKLNEAOKSL 134  
Db 732 FTVRNVRNTRFLPSNEKQNFQDLIPKEYEEVVKFEIETCTLYKQSEADLSNDPDLKTM 791  
Qy 135 LVSSQSFNNASGKLALDS--QLTNDFSEKSSY----- 165  
Db 792 IAQVSAVQKLEKSYKALRSYDVEEDFEFSEVVEQLPQYQGTENIKTIKEMIEDEGH 851  
Qy 166 -----FQSQVDKIRKEAYAGAAAGVAGPGLIISVSI-----NAGV 202  
Db 852 PESDFEKLQEIATFSSQLNATHKD-----VVDSPYINQLLKAQLNEAGA 896  
Qy 203 VEGKLIPELAKNKLKSVQNFPTTILSNTVKQANKDIDAAKLKLT---EIAAIGETKTET 259  
Db 897 VE-KFEKSIQKDPQIQKMYHTLKDLQVWTEIDVAQLKETSQNEI---QRQLQKEA 951  
Qy 260 TRFYVDYDDLMLSLKKAAMKMINTCNEYOKRHKGTLP 297  
Db 952 EEFGLSGDFLQSAM-----NEYQS--DKKTI 975

RESULT 45  
A90551  
conserved hypothetical protein MYPV\_3130 [imported] - Mycoplasma pulmonis (strain UAB CT  
C;Species: Mycoplasma pulmonis  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C;Accession: A90551  
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galissou, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
A;Reference number: A99512; MUID:21267165; PMID:11353084  
A;Accession: A90551  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2819 <KUR>  
A;Cross-references: UNIPROT:Q98Q98; GB:AL445566; PID:G14089727; PIDN:CAC13486.1; GSPDB:G  
A;Experimental source: strain UAB CTIP  
C;Genetics:  
A;Gene: MYPV\_3130  
A;Genetic code: SGC

Query Match 7.2%; Score 108.5; DB 2; Length 2819;  
Best Local Similarity 21.3%; Pred. No. 1.5e+02;  
Matches 77; Conservative 55; Mismatches 119; Indels 111; Gaps 20;

Qy 7 DKTVEVVVKNIAETADGALDLYNKYLDQVWPQTFDETIKELSRFKQBYSOAASVLVDIK 66  
Db 1820 DSIWNFVNESLWLE-KTDLIDKLDLW-----SSFNAAVSDKTKEIPNFS 1866  
Qy 67 TLL---MDSODKYFEATQV--YEWCGVATOLLAAAYILLFDEYNEKASAQ-----K 113  
Db 1867 SLLKPIMDQNDYPSLVKTLATINWKLMSQNTDAL-----KVALGQMLLRVPKNT 1917  
Qy 114 DILIKVLDDGTTKLN---EAQKSLVSSQSFNNASGKLLALDSQLTNDFFSEKSSYFOSQV 170  
Db 1918 DLNKLFDPLVQNLNFPIDQDKSKTNENLRNASKQILVF--VLQNE--DTLEFFKSLV 1973  
Qy 171 DK1--RKEAYAGAAGVAVGPFGLIISYSIAAGVVE-----CKLPEL--KNKLK 216  
Db 1974 NNIISNNQYKNA-----NNYDLIVGF-----IEKNKNLVTOYFKKIAPKIASDSMLK 2022  
Qy 217 SVQNFPTTSLNTVQKANDDAAKLKTTEIAA-----LGEIK--- 254  
Db 2023 TFVNGFAL--NLLQEQNLELSKDELSLVNSITSGAIDALSTSNIVEKILVFLAEVKKQN 2080  
Qy 255 TETETTRFYVDYDMLLS-----LLKE--AAKQMINTCNEYQKRHGKTKL 297  
Db 2081 PQSTNDLFAIQEILKSKDFSQNGLFEVLISILKSNILDSKINNKEYIQK---VKTL 2137  
Qy 298 FE 299  
Db 2138 FE 2139  
RESULT 46  
A:86248  
protein T23J18.9 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: A86248  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
C:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, H.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: A86248  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-636 <STO>  
A:Cross-references: UNIPROT:Q9LPZ2; GB:AE005172; NID:G6554201; PIDN:AAF16647.1; GSPDB:GN  
A:Gene: T23J18.9  
A:Map position: 1  
Query Match 7.2%; Score 108; DB 2; Length 636;  
Best Local Similarity 20.6%; Pred. No. 24;  
Matches 59; Conservative 39; Mismatches 88; Indels 100; Gaps 11;  
Qy 6 ADKTVEVVVKNIAETADGALDLYNKYLDQVWPQTFDETIKELSRFKQBYSOAASVLVDIK 56  
Db 412 SDRTFNVVKN---SASNAEENHAKHTIMVLPFAKSPVMYKTESLEVPKSVS-----HSL 463  
Qy 57 AASVLVGDITKLLMDSODKYFEATQVYEWCGVATOLLAAAYILLFDEYNEKASAQKDIL 116  
Db 464 HFSPL-----PETKQDFRE--GYALGMVTVYFGLLEKFKLEA----- 499  
Qy 117 IKVLDDGTTKLNKAEQKSL--VSSQSFN-----NAGSKLLAL-----DSQLT 156  
Db 500 ---DVPVSQNLNSKDSFSELEKGFNVTTPLSRIDKLLALKDROLTYIMEBLKGFDEMT 555  
Qy 157 NDFSEKSSYFOSQVDKIRKEAYAGAAGVAVGPFGLIISYSIAAGVVEGKLIPKLNK 216

Db 556 NEFSKAKQFDDMEQKIL-----EVKHKII 580  
Qy 217 SVQNFPTTSLNTVQKANDDAAKLKTTEIAAIGETETETTRF 262  
Db 581 ELQREAAALKEQAEKEQKDAAWKKICOMESCAKDLNVELEDVEF 626  
RESULT 47  
E97778  
endopeptidase La (EC 3.4.21.53) - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: E97778  
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro  
Science 293, 2093-2098, 2001  
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893  
A:Accession: E97778  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-779 <KUR>  
A:Cross-references: UNIPROT:Q92HZ1; GB:AE006914; PIDN:AAL03167.1; PID:gl5619715; GSPDB:G  
C:Gene: lon  
C:Superfamily: ATP-dependent Lon protease  
C:Keywords: Hydrolase; serine proteinase  
Query Match 7.2%; Score 108; DB 2; Length 779;  
Best Local Similarity 20.1%; Pred. No. 30;  
Matches 62; Conservative 61; Mismatches 102; Indels 84; Gaps 14;  
Qy 2 TEIVADKTEVVVKNIAETADGALDLYNKYLDQVWPQTFDETIKELSRFKQBYSOAASVL 61  
Db 71 TAILA-KIIQIVKLPNNTA-----KILIEAVA-RVKLSNKGEEAFENYIIPDEE 120  
Qy 62 VGDIKTL--LMDSODKYFEATQVYEWCGVATOLLAAAYILLFDEYNEKASAQ-----K 113  
Db 121 IFDVNNRSVLVDN-----AVOLFYSKAI-----NDKVNAILIETINK 158  
Qy 114 DI-----LIKVLDDGTTKLNKAEQKSLVSSQSFNNASGKLLALDSQLTNDFFSEK-- 163  
Db 159 EISNSTNFIDIIINILASHLITSLKAEQKLEETSPFKRITTVISMLNSNIVNSVETEQALQ 218  
Qy 164 SYFQSQVDKIRKEAYAGAAGVAVGPFGLIISYSIAAGVVEGKLIPKLNKLSVQNF 223  
Db 219 KRVRKQIEKTQDYLYLHEQMKAIQ-----KELEDKSELADIENKIK 260  
Qy 224 TLSNTVQKANDDAAKLKTTEIAAIGETETETTRFYVD-----YDDLMLSL 274  
Db 261 SLKLS-KEAKEKAEAEKLLRT-----MNQMSASGVTRNYLETLLSLPWGKYDNSKIDI- 314  
Qy 275 KEAAKMIN 283  
Db 315 -NQAEKILN 322  
RESULT 48  
T30336  
nuclear/mitotic apparatus protein - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T30336  
R:Merdes, A.; Ramyar, K.; Vechio, J.D.; Cleveland, D.W.  
Cell 87, 447-458, 1996  
A:Title: A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle assem  
A:Reference number: Z20828; MUID:97053784; PMID:8898198  
A:Accession: T30336  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2253 <MER>  
A:Cross-references: UNIPROT:P70012; EMBL:Y07624; NID:gl514670; PIDN:CAA68905.1; PID:gl51

|                       |        |                    |       |                                     |
|-----------------------|--------|--------------------|-------|-------------------------------------|
| Query Match           | 7.2%;  | Score 108;         | DB 2; | Length 2253;                        |
| Best Local Similarity | 19.3%; | Pred. No. 1.2e+02; |       |                                     |
| Matches               | 69;    | Conservative       | 60;   | Mismatches 146; Indels 72; Gaps 10; |

  

|    |      |   |      |
|----|------|---|------|
| QY | 6    | ADKTVVVKNAIETAGCALDLNNKYLDQVTPMQTFD-ETIKELSRFKQEYSQAASVLVGD | 64   |
| Db | 1132 | AEEEMAKLKSIVISEKSKRIEC-----LEQDIQNQRDLSCIEQHQSGLGSEQLQALIAI | 1187 |
| QY | 65   | I-----KTLLMDSODKYFEATQTVYEWCGVATQLAAVILLFDFEYNEKKASA-----Q  | 112  |
| Db | 1188 | LEKKCKEQEKELICEAQNKAAEAATLASEKASVSERQLEGIOALEIGKERACDLOKQ   | 1247 |
| QY | 113  | KOTILIKVLDGDTKLANEAKSLL-----VSSQSFNNAASKLLALDS-----QLTNDF   | 159  |
| Db | 1248 | LELSWAVEEEKETELQALKKELPHVKYQLEQSTSTFDSSGEALLYLSEAQRQQALTEA  | 1307 |
| QY | 160  | SEKSYFSQSQVDKIRKEAYAGAAGVVGAPFGIIISYIAAGV-----VEGKLIPELKNK  | 214  |
| Db | 1308 | KEAQEQYQKEIEMKNKEVNSLQAE-----IKILSSKVTTNEEVSDPFOURLLKETS    | 1360 |
| QY | 215  | LKSVQNFPTLTNTVQANKIDAAKLLTETAAGEIKTKETETTRFYDYDDMLSL        | 274  |
| Db | 1361 | AK-----LEEKMQLHMELEASPFLERKNCADCLTTEAQNKLGEADOQORMAVDSL     | 1412 |
| QY | 275  | -----KEAAKMINTCNEYQKRGHKTLFE                                | 299  |
| Db | 1413 | QQKLSSKAETNHTYQEIQAQWNCAKEQQICSLQONLUKSNOSLLIE              | 1459 |

  

RESULT 49

T18281

hypothetical protein D1 - slime mold (Dictyostelium discoideum)

C:Species: Dictyostelium discoideum

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T18281

R:Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh Genetics 148, 1117-1125, 1998

A:title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1 A:Reference number: Z14684; MUID:98198836; PMID:9539429

A:Accession: T18281

A>Status: preliminary; translated from GB/EMBL/DDBB

A:Molecule type: DNA

A:Residues: 1-1474 <RIE>

A:Cross-references: UNIPROT:Q33870; EMBL:U00796; NID:g2702254; PID:g2702256; PIDD:AAC186 C:Genetics:

A:Introns: 33/2

|                       |        |   |                 |                     |
|-----------------------|--------|---|-----------------|---------------------|
| Query Match           | 7.1%   | Score 107;  | DB 2;           | Length 1474;        |
| Best Local Similarity | 22.0%; | Pred. No. 79;   |                 |                     |
| Matches               | 79;    | Conservative 48;  | Mismatches 129; | Indels 90; Gaps 16; |
| Qy                    | 18     | ETADGALDL-----YNKYLQOVIPWQTFDETIKELSRPKQEYSQAASVL--VGDIKT 67    |                 |                     |
| Db                    | 777    | ESQGSNDIBIQKLNQSKYISTI---NDRDSTIKSLQALINELASSIFKLNQSSIKD 833    |                 |                     |
| Qy                    | 68     | LMDSQDKYPEATQTVYEGWGATQLLAAAYILLFDEYNEKKASAQ--KQILIKVLDDGIT 125 |                 |                     |
| Db                    | 834    | TLFNSQALLIEKQKSN---NMVLEKSEQEYKILLDQIEKKNMNVNSYEIKTKDETIE 890   |                 |                     |
| Qy                    | 126    | KLNEAQKSLVSSQSFNNASGKL-LALDSQ---LTNDFSEKSYF-----SQVQDKI- 173    |                 |                     |
| Db                    | 891    | VLNQTILNCTNESNTIETVKLHLLENKIALLLNEISSKQLYFDTISGTQYNYEVE 950     |                 |                     |
| Qy                    | 174    | -----RKEYAGAAAGVAGPGLIISYISIAAGVVEGKLIPELKNKLKSVQNF--TL 225     |                 |                     |
| Db                    | 951    | NCFSEKE-----LQVSY--AARGVLKOKNDQOINAELETKNKFDPNL 993             |                 |                     |
| Qy                    | 226    | SNTVKQANK-----DIDAAKLTLTTEIAAIGIKETETT-----REYVDYD- 267         |                 |                     |
| Db                    | 994    | LNLSIKTNEHKSQINDLNTQNYGLEIEBSRIIQLETTPTVSNQIQTQPAPEYSYKHE 1053  |                 |                     |
| Qy                    | 268    | -----DLMLSLLKEAAK-----KWINTCNEYQKRHGKKT 296                     |                 |                     |

```

Db      1054 ILKRDLSLKLNERAKVYKYISTSVPELTSGINTNNRNSKNNTT 1099
          |::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 50
TI3329
hypothetical protein 40 - Streptococcus phage phi-O1205
C:Species: Streptococcus phage phi-O1205
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C:Accession: TI3329
R:Stanley, E.; Fitzgerald, G.F.; Le Marrec, C.; Fayard, B.; van Sinderen, D.
Microbiology 143, 3417-3429, 1997
A:Title: Sequence analysis and characterization of phi O1205, a temperate bacteriophage
A:Reference number: Z17654; MUID:98049466; PMID:9387220
A:Accession: TI3329
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1517 -STA>
A:Cross-references: UNIPROT:O34071; EMBL:U89374; NID:g2444080; PID:g2444119; PIDN:AAC795F
A:Experimental source: host Streptococcus thermophilus strain CNRZ1205

Query Match       7.1%; Score 107; DB 2; Length 1517;
Best Local Similarity 20.0%; Pred. No. 82;
Matches 59; Conservative 53; Mismatches 123; Indels 60; Gaps 11;

QY    3   EVADKTVEVVK-----NAETPADGALDLYNKYLQGVIPWQTF-----DETIKELSRP 50
      ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db    246 ETVSNIKLPVFPEKLSAKAIRGIEIGIIDAFSKIDDSKI--QSPANNLSKGIDKTVKDISQT 303
      ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY    51 KQSY-----SQASVLVG-DIKTLMLDSQDKYPEATQTVVYECG 88
      ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db    304 VQSFWEGFSNTSAIKGTNAPRYVASQISLAFKGIDFRNLFKGLGVFGDIAGVISRTLT 363
      ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY    89 VATQLLAAYILLFDYNEKKA--SAQKD--ILIKVLDDGITKL---NEAQKSLLVSSQSIF 141
      ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db    364 IATKSVSNFISSPADTGAFRAFKTALEDVVVVVKLGSSLADVFNSENSEMQTTIILGTAF 423
      ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY    142 NNASGKKLLADSQLTNDPSEKSSVPQSOVDKIRKEAYAGAAAGVAVGPGLI----LSYSI 198
      ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db    424 GT-----LAKVWSQAASAIAKFPVSSIPKSVLNGITSGIIAIAAGFTYAKVGISA 472
      ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY    199 AAGVGEGLIPELKNKLKSVQNFFTTLSNTVKQANKDIDAACKLKTTEATAIGEI 253
      ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db    473 VGALKGL-----DFINSLNPFKKFGKDAABEGTEQAASRSKSTITQLEFSGISNV 524
      ||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

Search completed: January 5, 2005, 10:59:47
Job time : 21.4424 secs

```

Search completed: January 5, 2005, 10:59:47  
Job time : 21.4424 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:42:38 ; Search time 63.7089 Seconds  
(without alignments)  
1717.381 Million cell updates/sec

Title: US-09-993-292B-2  
Perfect score: 1515  
Sequence: 1 MTSIFAEQTVVVKSALETA.....NEVQRHGKTLFEVPDVS 305

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database : A Genesep23Sep04:\*

- 1: Genesep21980s:\*
- 2: Genesep21990s:\*
- 3: Genesep22000s:\*
- 4: Genesep22010s:\*
- 5: Genesep22020s:\*
- 6: Genesep22030s:\*
- 7: Genesep22040s:\*
- 8: Genesep22050s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 1515  | 100.0       | 305    | 6     | ABG73265 Salmonell |
| 2          | 1515  | 100.0       | 749    | 6     | ABG73266 Salmonell |
| 3          | 1390  | 91.7        | 309    | 2     | AAR67645 Haemolysi |
| 4          | 1079  | 71.2        | 1035   | 4     | ABG25221 Novel hum |
| 5          | 121.5 | 8.0         | 1036   | 7     | ADF07451 Bacterial |
| 6          | 118.5 | 7.8         | 2681   | 6     | ABJ19025 Pathogen  |
| 7          | 117.5 | 7.8         | 582    | 6     | ABM68103 Phototrab |
| 8          | 114.5 | 7.6         | 588    | 4     | ABM71280 Drosophil |
| 9          | 114   | 7.5         | 515    | 7     | ADC95690 E. faeciu |
| 10         | 112.5 | 7.4         | 1948   | 5     | ABP73774 Candida a |
| 11         | 112   | 7.4         | 1189   | 6     | ABU42996 Protein e |
| 12         | 111.5 | 7.4         | 1158   | 6     | ABU28937 Protein e |
| 13         | 111   | 7.3         | 941    | 6     | ABU18419 Protein e |
| 14         | 111   | 7.3         | 1231   | 7     | ADC95523 E. faeciu |
| 15         | 110.5 | 7.3         | 476    | 2     | AAR43563 Hyalurona |
| 16         | 110   | 7.3         | 951    | 4     | ABU53070 Intracell |
| 17         | 110   | 7.3         | 953    | 4     | ABU53071 Intracell |
| 18         | 110   | 7.3         | 961    | 4     | ABU53077 Intracell |
| 19         | 110   | 7.3         | 2228   | 7     | ABR61599 Human gol |
| 20         | 110   | 7.3         | 2230   | 6     | ABU07445 Protein d |
| 21         | 110   | 7.3         | 2230   | 7     | ABR61600 Human gol |
| 22         | 110   | 7.3         | 2250   | 7     | ABR61601 Human gol |
| 23         | 110   | 7.3         | 2252   | 7     | ABR61602 Human gol |
| 24         | 109.5 | 7.2         | 742    | 6     | ABU25599 Protein e |
| 25         | 109   | 7.2         | 739    | 4     | AAB96493 Putative  |

|    |       |     |      |   |                    |
|----|-------|-----|------|---|--------------------|
| 26 | 109   | 7.2 | 929  | 4 | ABU53079 Intracell |
| 27 | 108.5 | 7.2 | 1048 | 4 | ABBS9245 Drosophil |
| 28 | 108.5 | 7.2 | 1066 | 4 | AAG67418 Amino aci |
| 29 | 108.5 | 7.2 | 1066 | 6 | ABG72693 Fruitfly  |
| 30 | 108.5 | 7.2 | 1066 | 7 | ADG98862 Fruit fly |
| 31 | 108.5 | 7.2 | 3692 | 6 | ABU43311 Protein e |
| 32 | 108.5 | 7.2 | 3696 | 5 | ABP40235 Staphyloc |
| 33 | 107.5 | 7.1 | 1292 | 5 | ABP77986 Amino aci |
| 34 | 107.5 | 7.1 | 1875 | 6 | ABR53560 Protein s |
| 35 | 107.5 | 7.1 | 1875 | 7 | ADK64380 Disease t |
| 36 | 107   | 7.1 | 1211 | 5 | ABP39975 Staphyloc |
| 37 | 106.5 | 7.0 | 927  | 5 | ABP49720 Listeria  |
| 38 | 106.5 | 7.0 | 1312 | 2 | AAW22775 Human RAD |
| 39 | 106.5 | 7.0 | 1312 | 2 | AAW71295 Human hom |
| 40 | 106.5 | 7.0 | 2346 | 4 | ABM63519 Drosophil |
| 41 | 106.5 | 7.0 | 4734 | 8 | ADH62807 Lactobaci |
| 42 | 106.5 | 7.0 | 7201 | 4 | ABM71136 Drosophil |
| 43 | 106   | 7.0 | 685  | 4 | ABM64018 Drosophil |
| 44 | 106   | 7.0 | 815  | 6 | ADA32997 Acinetoba |
| 45 | 106   | 7.0 | 863  | 6 | ADB08930 Alloiococ |
| 46 | 106   | 7.0 | 885  | 4 | AAW82283 S. epider |
| 47 | 106   | 7.0 | 2835 | 5 | AAU75100 Novel sig |
| 48 | 105.5 | 7.0 | 1318 | 5 | ABP77985 Amino aci |
| 49 | 105.5 | 7.0 | 1318 | 7 | ADJ68860 Human hea |
| 50 | 105.5 | 7.0 | 1318 | 8 | ADJ66509 RAD50 hom |

## ALIGNMENTS

## RESULT 1

ABG73265  
ID ABG73265 standard; protein; 305 AA.

XX ABG73265;

DT 17-APR-2003 (first entry)

XX Salmonella typhi cytolysin A (ClyA) protein.

DE Protein expression vector; expression cassette; export protein;  
XX transformed host cell; bacterial cell; immune response; HlyE family;  
KW cytolysin A; ClyA; recombinant protein; bioreactor; bacterial growth;  
KW live vector immunogenic composition; animal feed; animal vaccination;  
KW food industry; nutritional supplement; biomediation; waste disposal;  
KW waste treatment; host bacterium.

OS Salmonella typhi.

XX US2002146430-A1.

PN 10-OCT-2002.

PD 23-NOV-2001; 2001US-00993292.

PP 22-NOV-2000; 2000US-0252516P.

XX (GALE/J) GALEN J E.

PI Galen JE;

XX WPI; 2003-228013/22.

DR N-PSDB; ABX15174.

XX Novel system for expressing protein of interest, has expression vector  
PT with cassette having export protein coding sequence fused to protein  
PT coding sequence, host cell transformed with vector, and culturing  
PT environment.

XX Claim 5; Page 18; 33pp; English.

XX The present invention relates to a system for expressing a protein of  
CC interest. The system comprises an expression vector having an expression

CC cassette, where the expression cassette comprises an export protein  
 CC coding sequence genetically fused to a DNA sequence encoding the protein  
 CC of interest, a host cell transformed with the expression vector, and a  
 CC culturing environment for the transformed host cell, where the expression  
 CC cassette expresses the export protein-protein of interest fusion protein  
 CC which is exported out of the host cell. The system is useful for  
 CC expressing a gene in a bacterial cell, by providing an expression vector  
 CC to a population of untransformed bacterial cells, and expressing the  
 CC expression cassette, such that the export protein-protein of interest  
 CC fusion protein is produced and exported into the culture medium. The  
 CC protein of interest is preferably an antigen and this method is useful  
 CC for eliciting an immune response from a host against the fusion protein.  
 CC The export protein may be a member of the HlyE family such as cytolyisin A  
 CC (ClyA). The system is useful for the production of recombinant proteins  
 CC inside a bacterial host cell, in a bio-reactor, and various devices that  
 CC facilitate the growth of bacteria. The system is also useful for  
 CC providing an animal antigenic material against which an immune response  
 CC may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can  
 CC be used as a live vector immunogenic composition capable of facilitating  
 CC the generation of an immune response in an animal. The system is also  
 CC useful for preparing immunogenic compositions for vaccinating animals,  
 CC and for use in the food, nutritional supplement, animal feed,  
 CC biomediation, waste disposal, and waste treatment industries. The system  
 CC is very advantageous for protein expression, as proteins that are toxic  
 CC to host bacterium can also be expressed. A population of recombinant host  
 CC cells can be transformed with the expression vector, and then the  
 CC population of bacterial host cells can be maintained in culture and used  
 CC to produce proteins without having to harvest and lyse the host cells.  
 CC The present represents Salmonella typhi ClyA protein

XX Sequence 305 AA;

Query Match 100.0%; Score 1515; DB 6; Length 305;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-116;  
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTSIFAEQTVVEVKSATETADGALDLYNKYLDQVTPWKTDETTIKELSRFKQEYSQASV 60  
 Db 1 MTSIFAEQTVVEVKSATETADGALDLYNKYLDQVTPWKTDETTIKELSRFKQEYSQASV 60

Qy 61 LVGDIKLLMDSQKPYEATQTVYEWGCVTQLLSAVILLFDEYNEKKAQKDILIRIL 120  
 Db 61 LVGDIKLLMDSQKPYEATQTVYEWGCVTQLLSAVILLFDEYNEKKAQKDILIRIL 120

Qy 121 DDGVKLENAQKSLTSSQSFNNAAGKLLALDSQLTNDFSEKSSYFQSDVRIRKEAYAG 180  
 Db 121 DDGVKLENAQKSLTSSQSFNNAAGKLLALDSQLTNDFSEKSSYFQSDVRIRKEAYAG 180

Qy 181 AAAGIVAGPGLIITSYIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVQANKDIDAAK 240  
 Db 181 AAAGIVAGPGLIITSYIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVQANKDIDAAK 240

Qy 241 LKLAETAAIGEIKTETTRFYVDYDMLSLKGAQKXMTNCNVOYQRHGKTLFEV 300  
 Db 241 LKLAETAAIGEIKTETTRFYVDYDMLSLKGAQKXMTNCNVOYQRHGKTLFEV 300

Qy 301 PDVAS 305  
 Db 301 PDVAS 305

RESULT 2

ABG73266

ID ABG73266 standard; protein; 749 AA.

XX AC ABG73266;

XX DT 17-APR-2003 (first entry)

XX Salmonella typhi ClyA-Bacillus subtilis sacB fusion protein.

XX Protein expression vector; expression cassette; export protein;

XX transformed host cell; bacterial cell; immune response; HlyE family;

KW cytolyisin A; ClyA; recombinant protein; bio-reactor; bacterial growth;  
 KW live vector immunogenic composition; animal feed; animal vaccination;  
 KW food industry; nutritional supplement; biomediation; waste disposal;  
 KW waste treatment; host bacterium; sacB; mutant; mutein;  
 KW lethal levansucrase.  
 XX Salmonella typhi.  
 OS Bacillus subtilis.  
 OS Synthetic.  
 OS Chimeric.  
 XX US2002146430-A1.  
 PN 10-OCT-2002.  
 XX 23-NOV-2001; 2001US-00993292.  
 XX 22-NOV-2000; 2000US-0252516P.  
 XX (GALE/) GALEN J E.  
 XX Galen JE;  
 DR WPI; 2003-228013/22.  
 DR N-PSDB; ABX15191.  
 XX Novel system for expressing protein of interest, has expression vector  
 PT with cassette having export protein coding sequence fused to protein  
 PT coding sequence, host cell transformed with vector, and culturing  
 PT environment.  
 XX Example 3; Page 25-28; 33pp; English.

The present invention relates to a system for expressing a protein of interest. The system comprises an expression vector having an expression cassette, where the expression cassette comprises an export protein coding sequence genetically fused to a DNA sequence encoding the protein of interest, a host cell transformed with the expression vector, and a culturing environment for the transformed host cell, where the expression cassette expresses the export protein-protein of interest fusion protein which is exported out of the host cell. The system is useful for expressing a gene in a bacterial cell, by providing an expression vector to a population of untransformed bacterial cells, and expressing the expression cassette, such that the export protein-protein of interest fusion protein is produced and exported into the culture medium. The protein of interest is preferably an antigen and this method is useful for eliciting an immune response from a host against the fusion protein. The export protein may be a member of the HlyE family such as cytolyisin A (ClyA). The system is useful for the production of recombinant proteins inside a bacterial host cell, in a bio-reactor, and various devices that facilitate the growth of bacteria. The system is also useful for providing an animal antigenic material against which an immune response may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can be used as a live vector immunogenic composition capable of facilitating the generation of an immune response in an animal. The system is also useful for preparing immunogenic compositions for vaccinating animals, and for use in the food, nutritional supplement, animal feed, biomediation, waste disposal, and waste treatment industries. The system is very advantageous for protein expression, as proteins that are toxic to host bacterium can also be expressed. A population of recombinant host cells can be transformed with the expression vector, and then the population of bacterial host cells can be maintained in culture and used to produce proteins without having to harvest and lyse the host cells. The present represents Salmonella typhi ClyA as a fusion partner to export heterologous antigens out of S. typhi

Sequence 749 AA;

Query Match

Best Local Similarity 100.0%; Score 1515; DB 6; Length 749;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSIFABQTVVVKSAIETADGALDLNKKYLDQVVPWKTFDETIKELSRFKQEYSQASV 60  
DB 1 MTSIFABQTVVVKSAIETADGALDLNKKYLDQVVPWKTFDETIKELSRFKQEYSQASV 60  
QY 61 LVGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQDILIRIL 120  
DB 61 LVGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQDILIRIL 120  
QY 121 DDGVKCLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAG 180  
DB 121 DDGVKCLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAG 180  
QY 181 AAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTLSATVQANKDIDAAK 240  
DB 181 AAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTLSATVQANKDIDAAK 240  
QY 241 LKLAETAAIGEIKTETETTRFYVDYDDLMLSLKGAAKKMWINTCNEYQORHGKKTLEFV 300  
DB 241 LKLAETAAIGEIKTETETTRFYVDYDDLMLSLKGAAKKMWINTCNEYQORHGKKTLEFV 300  
QY 301 PDVAS 305  
DB 301 PDVAS 305

RESULT 3  
AAR67645  
ID AAR67645 standard; protein; 309 AA.  
XX  
AC AAR67645;  
XX  
DT 25-MAR-2003 (revised)  
DT 06-JUL-1995 (first entry)  
XX  
DE Haemolysin regulator.  
XX  
KW Haemolysin regulator; tuberculosis; vaccine; therapy.  
XX  
OS Escherichia coli.  
XX  
PN WO9428137-A1.  
XX  
PD 08-DEC-1994.  
XX  
PF 24-MAY-1994; 94WO-US005869.  
XX  
PR 24-MAY-1993; 93US-00066830.  
XX  
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI King CH, Shinnick TM, Sathish M;  
XX  
DR WPI; 1995-022809/03.  
DR P-PSDB; AAQ75857.  
XX  
PT New nucleic acid encoding haemolytic regulator of E coli - and derived  
PT vectors, proteins etc used to transform Mycobacterium cells to produce  
PT antituberculosis vaccines providing improved immune response.  
PS Claim 9; Page 39-40; 46pp; English.  
XX  
CC The sequence represents the E. coli haemolysin regulator which may be  
CC used as an immunogen in a vaccine composition against tuberculosis  
CC (Mycobacterium tuberculosis). (Updated on 25-MAR-2003 to correct PN  
CC field.) (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-  
CC 2003 to correct PA field.)  
SQ Sequence 309 AA;

Query Match 91.78; Score 1390; DB 2; Length 309;  
Best Local Similarity 91.08; Pred. No. 1.7e-105;  
Matches 274; Conservative 15; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTSIFABQTVVVKSAIETADGALDLNKKYLDQVVPWKTFDETIKELSRFKQEYSQASV 60  
DB 1 MTSIFABQTVVVKSAIETADGALDLNKKYLDQVVPWKTFDETIKELSRFKQEYSQASV 60  
QY 61 LVGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQDILIRIL 120  
DB 61 LVGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQDILIRIL 120  
QY 121 DDGVKCLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAG 180  
DB 121 DDGVKCLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAG 180  
QY 181 AAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTLSATVQANKDIDAAK 240  
DB 181 AAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTLSATVQANKDIDAAK 240  
QY 241 LKLAETAAIGEIKTETETTRFYVDYDDLMLSLKGAAKKMWINTCNEYQORHGKKTLEFV 300  
DB 241 LKLAETAAIGEIKTETETTRFYVDYDDLMLSLKGAAKKMWINTCNEYQORHGKKTLEFV 300  
QY 301 P 301  
DB 301 P 301  
RESULT 4  
ABG25221  
ID ABG25221 standard; protein; 1035 AA.  
XX  
AC ABG25221;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #25212.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS89408.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 55580; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food





XX New isolated nucleic acid derived from *Enterococcus faecium* encoding an  
PT *Enterococcus faecium* polypeptide useful for detection, prevention and  
PT *Enterococcus faecium* polypeptide useful for detection, prevention and  
PT treatment of a pathological condition resulting from a bacterial  
PT infection.

XX Example 1; SEQ ID NO 5317; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from  
CC *Enterococcus faecium* encoding an *Enterococcus faecium* polypeptide having  
CC one of 10 fully defined sequences given in the (or comprising 40  
CC sequential nucleotides chosen from any of the nucleic acids, its  
CC complement or sequences hybridising to it). Also included are a  
CC recombinant vector comprising the nucleic acid operably linked to  
CC transcription regulatory element, a cell comprising the vector and a  
CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
CC The nucleic acids is useful for diagnosing pathological conditions  
CC resulting from *E. faecium* bacterial infection (e.g. urinary tract  
CC infection, bacteraemia, endocarditis, wounds and abdominal pelvic  
CC infection) and for screening drugs such as agonists and antagonists. The  
CC nucleic acid is useful for recombinant production of *Candida albicans*.











XX 01-JAN-2004 (first entry)

XX E. faecium protein sequence SEQ ID 5150.

XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;

XX abdominal-pelvic infection.

XX Enterococcus faecium.

XX US6583275-B1.

XX 24-JUN-2003.

XX 30-JUN-1998; 98US-00107532.

XX 02-JUL-1997; 97US-0051571P.

XX 14-MAY-1998; 98US-0085598P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2003-799836/75.

XX N-PSDB; ADC91869.

XX New isolated nucleic acid derived from Enterococcus faecium encoding an

XX Enterococcus faecium polypeptide useful for detection, prevention and

XX treatment of a pathological condition resulting from a bacterial

XX infection.

XX Example 1; SEQ ID NO 5150; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from

XX Enterococcus faecium encoding an Enterococcus faecium polypeptide having

XX one of 10 fully defined sequences given in the (or comprising 40

XX sequential nucleotides chosen from any of the nucleic acids, its

XX complement or sequences hybridising to it). Also included are a

XX recombinant vector comprising the nucleic acid operably linked to

XX transcription regulatory element, a cell comprising the vector and a

XX single-stranded probe comprising the nucleic acid. The nucleic acids are

XX chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.

XX The nucleic acids are useful for diagnosing pathological conditions

XX resulting from E. faecium bacterial infection (e.g. urinary tract

XX infection, bacteraemia, endocarditis, wounds and abdominal-pelvic

XX infection) and for screening drugs such as agonists and antagonists. The

XX nucleic acid is useful for recombinant production of Candida albicans -

XX derived peptides or antisense polypeptides. Pharmaceutical compositions

XX and vaccines containing the nucleic acid are useful for preventing or

XX treating Enterococcus faecium infections. The present sequence represents

XX one if the disclosed E. faecium proteins.

XX

XX Sequence 1231 AA;

XX Query Match 7.3%; Score 111; DB 7; Length 1231;

XX Best Local Similarity 23.3%; Pred. No. 6.1;

XX Matches 78; Conservative

XX 6 AEQTVVVKGAETADGALDLYNK-----YLDQVTPWKTDE-----TIKELS 48

XX 269 AEEVIVSMGVAQTEGTIYLVNANGKGTGFLNIHLRYFPFVTFQKLPHVTKTIADVLD 328

XX 49 RFKQESQASVLVDIKVILMSQ-----DKYFEATQVYVWCVGTQLLSAVILLF 101

XX 329 RSKFPGAGGEPPLL-DVQSALYDSELPAVIGRGYGLGSK-----DVTDPQISA---VF 378

XX 102 DEYNEXKASAKOILILILDGVKKNEAQKSL-LTSSQSFN-----NAGSKL 148

XX 379 DELKKDPSIRKRRFTIGVDVTYQSLPEKESLDLTPFQAKFWGFGSDGTVGANKSA 438

XX 149 LALDSQNLNDFSEKSSVFSQ-----VDRIR-----KEAYAGAAAGIVA--GPFGLII 194

XX 439 IKIICDHTDKYAQGYFYDYSKSGGLTVSHLRFQDTPIRSAVLVEHADLVACHPT-AYLH 497

QY 195 SYSTAGVIRGKLIPELNNRLKTVQFFTSLSATVQ--ANKDI---DAAKUKLATEIAA 249

DB 498 SYDLVKGKPGGIF--LLNTLWSDEQLETHLPKUKRYLAENNIRFYTNAMRLAQEVGL 555

QY 250 IGEIKTETETTRF-----YVDYDDMLSLKGAACK 280

DB 556 GRRINTAMETAFFKLADIIPFDE-VLPFLKBEALK 589

RESULT 15

AAR43563

ID AAR43563 standard; protein; 476 AA.

XX AAR43563;

AC AAR43563;

XX 25-MAR-2003 (revised)

DT 09-JAN-2003 (revised)

DT 05-APR-1994 (first entry)

XX Hyaluronan receptor.

DE Hyaluronan receptor.

XX Hyaluronan binding protein; HA; RHAMM; mediated motility; wound; healing;

XX diagnosis; treatment; cell locomotion; tumour invasion; birth defects;

XX inflammatory disorder; Alzheimer's disease; dementia;

XX Parkinson's disease; Huntington's disease; AIDS; diabetes; auto;

XX immune diseases; corneal dysplasia; hypertrophy; surgery; burns; strokes;

XX multiple sclerosis; depression; schizophrenia; CNU; contraception;

XX in vitro fertilisation; embryo development.

XX Rattus sp.

XX WO9321312-A1.

PN 28-OCT-1993.

PD 13-APR-1993; 93WO-CA000158.

PF 09-APR-1992; 92GB-00007949.

PR (UYMA-) UNIV MANITOBA.

PA (MANI-) MANITOBA CANCER TREATMENT & RES FOUND.

PI Turley EA;

XX WPI; 1993-351722/44.

XX N-PSDB; AAQ51212.

XX DNA encoding hyaluronan receptor - used to produce proteins and

XX antibodies for alteration of cell locomotion.

XX Claim 7; Fig 23; 88pp; English.

XX The sequence is that encoded by a cDNA clone encoding the hyaluronan

XX receptor (HARC). The sequence was obtd. by screening a 3T3 library in

XX lambda gt11 with antibodies to HARC. A clone of 1.9 kb was obtained and

XX used to rescreen the library to obtain the full length, 2.9 kb clone. HA

XX is down regulated in stationary normal cells and is only expressed in

XX situations where cell motility is desired, e.g. in wound healing, in

XX response to growth factors and in chemotaxis by white blood cells. HA may

XX be used for diagnosis and treatment of diseases involving cell

XX locomotion, e.g. tumour invasion, birth defects, acute and chronic

XX inflammatory disorders, Alzheimer's and other forms of dementia, AIDS,

XX diabetes, autoimmune diseases, corneal dysplasias and hypertrophies,

XX burns, surgical incisions and adhesions, strokes, multiple sclerosis,

XX depression/schizophrenia related to neuronal growth and pain states

XX involving nerve sprouting; also in CNJ and spinal cord regeneration,

XX contraception, in vitro fertilisation and embryo development. See also

XX AAR46548-51. (Updated on 09-JAN-2003 to add missing OS field.) (Updated

XX on 25-MAR-2003 to correct PN field.)

XX Sequence 476 AA;

Query Match 7.3%; Score 110.5; DB 2; Length 476;  
Best Local Similarity 20.4%; Pred. No. 2;  
Matches 68; Conservative 73; Mismatches 125; Indels 67; Gaps 18;  
QY 1 MTSIFARQTVVVKSAIETADGALDLYNKYLDOVIPWKTFDETIKELSRFKQYEQEASV 60  
DB 99 LDNLLREKEVELEKHIAHQQAIIAQEKYIDTA--QSLRVVVAQLESVQEKYNDTAQS 155  
QY 61 LVGDIKVLMDSQDKYFEATQTVYEWGCVVTQLLSAYILLFDEYNEKASAKQDKILIRIL 120  
DB 156 L-RDVTAQLESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQLSLRDVTAQ-L 205  
QY 121 DGGVKLNEAKSILLTSS---QSPNNSAGKLLALDLSOLTN-DFSEKSSYFQSQVDRIKE 176  
DB 206 ESQVEKYNDAQSLRDVSAQLESYKSTLKEIE-DLKGENTLQEKVAMAEKSVEDVQQQ 264  
QY 177 AYAGAAAGIVAGPFLGLIISYIAAGVIEGKLIPELNRR-----LKTQVQNF--TSL 225  
DB 265 ILTAESTN-----QEVA-----RMVQDLQNRSTLKEEIEKITSSFLEKITDL 307  
QY 226 SATVQANKOI-----DAKCLKATEI-----AAGIKETETETTRFYVD-YD 267  
DB 308 KNQLRQODEDFRQLEEKGRTAENVMTELTMEINKWALLYELYEKTQPFQOQLDAFE 367  
QY 268 DLMLSLIK--GAAKMINTCNE-YQQRHGKTL 297  
DB 368 AEQALLNEHGATQEQUNKIRDSYAQLLGHQNL 400

RESULT 16  
ABU53070  
ID ABU53070 standard; protein; 951 AA.  
AC ABU53070;  
XX  
DT 15-APR-2003 (first entry)  
XX  
DE Intracellular trafficking-associated DKFZphtes3\_lg13 homologue #3.  
XX  
KW Human; gene therapy; vaccine; disease treatment; detection.  
XX  
OS Homo sapiens.  
XX  
PN W0200112659-A2.  
XX  
PD 22-FEB-2001.  
XX  
PF 18-AUG-2000; 2000WO-IB001496.  
XX  
PR 18-AUG-1999; 99US-0149499P.  
PR 28-SEP-1999; 99US-0156503P.  
XX  
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX  
PI Wiemann S;  
XX  
DR WPI; 2001-327840/34.  
XX  
XX Nucleic acids having the sequences of clones isolated from libraries of  
XX different human tissues, useful in recombinant DNA methodologies.

PS Example III; Page 667; 1095pp; English.  
XX  
CC This invention describes novel polynucleotides and polypeptides isolated  
CC from human cDNA libraries which can be used for gene therapy or in  
CC vaccines. The polynucleotides of the invention and antibodies encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate polypeptide expression. The products of the  
CC invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of  
CC the invention may also be used as diagnostic agents for detecting the  
CC presence of polypeptides in samples. This sequence represents a homologue  
CC of a polypeptide described in the disclosure of the invention

XX Sequence 951 AA;  
SQ Query Match 7.3%; Score 110; DB 4; Length 951;  
Best Local Similarity 20.6%; Pred. No. 5.3;  
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;  
QY 7 EQTVVVVKSAIETADGALDLYNKYLDOVIPW-KTFDETIKELSRFKQY-----SOEA 58  
DB 394 EKQMEKVKQAKEMQETLK--KKLLDQEAKLKKELENTALELSQEKQFNAKMLEMAQAN 451  
QY 59 SVLVGDIKVLMDSQDKYFEATQTVY--EWCQGVVT-----QLLSAYILLFDEYN 105  
DB 452 SAGISDAVSELETNQEQIESLTVHRRELNDVISWEKLNQQAELQEIHIQLOEKE 511  
QY 106 EKKASAKQDILI-----RILDDGVK--LNEAKSLTSSQSFNNASKLL 149  
DB 512 QEVAELKQKILLPGCEKEEMNKETWLKBEQVQDITLNELOQLKQKSAHVNSLAQDET 571  
QY 150 ALDSQL-----TNDSEKSSYFQSQVDRIKEAYAGAAAGIVAGPFLGLIISYIAAGVI 203  
DB 572 KKAHLEKLEVDLNLKSKENTFLOQLVELKMLAE----- 607  
QY 204 EGKLIPELNRLKTVQNFFTSLSATVKQANKIDAAKL---KLATEIAAIGEI---KTET 257  
DB 608 DRKVSLETSKLTDEEFQSLKSSHEKSNKSLSEDKSLEFKLSELAIQIDICCKTE- 666  
QY 258 ETTRFYVDYDDLMLSLKGAAKKMIN 283  
DB 667 -----ALLEAKTNELIN 678

RESULT 17  
ABU53071  
ID ABU53071 standard; protein; 953 AA.  
AC ABU53071;  
XX  
DT 15-APR-2003 (first entry)  
XX  
DE Intracellular trafficking-associated DKFZphtes3\_lg13 homologue #4.  
XX  
KW Human; gene therapy; vaccine; disease treatment; detection.  
XX  
OS Homo sapiens.  
XX  
PN W0200112659-A2.  
XX  
PD 22-FEB-2001.  
XX  
PF 18-AUG-2000; 2000WO-IB001496.  
XX  
PR 18-AUG-1999; 99US-0149499P.  
PR 28-SEP-1999; 99US-0156503P.  
XX  
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX  
PI Wiemann S;  
XX  
DR WPI; 2001-327840/34.  
XX  
XX Nucleic acids having the sequences of clones isolated from libraries of  
XX different human tissues, useful in recombinant DNA methodologies.

PS Example III; Page 668; 1095pp; English.  
XX  
CC This invention describes novel polynucleotides and polypeptides isolated  
CC from human cDNA libraries which can be used for gene therapy or in  
CC vaccines. The polynucleotides of the invention and antibodies encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate polypeptide expression. The products of the  
CC invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of  
CC the invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of

CC the invention may also be used as diagnostic agents for detecting the  
CC presence of polypeptides in samples. This sequence represents a homologue  
CC of a polypeptide described in the disclosure of the invention  
XX  
SQ Sequence 953 AA;  
Query Match 7.3%; Score 110; DB 4; Length 953;  
Best Local Similarity 20.6%; Pred. No. 5.3;  
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;  
QY 7 EOTVEVVKSAIETADGALDLYNKLDQVTPW-KTFDETIKELSRFKQY-----SQEA 58  
Db 484 EKKMEVKQAKEMQETLK--KLLDQEAFLKKELENTALELSQKEQFNKMLEMAQAN 541  
QY 59 SVLVGDIKVLMDSDQKYFRATQVY--EMCGVVT-----QLLSAYILLDFEYN 105  
Db 542 SAGISDAVSRLTQKEQIESLTVHRRELNDVISWEKLNQQAELQEIHIQLOEKE 601  
QY 106 EKKASAQKDILI-----RILDDGVKK--LNEAQSKLLTSSQSFNNASGKLL 149  
Db 602 QEVAELKQKILLFGCEKEEMNKETWLKEGVKQDITLNEQLQKSAHVNSLAQDET 661  
QY 150 ALDSQL-----TNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPFGLLIISYIAAGVI 203  
Db 662 KLAHLEKLEVDLNLKSLKENTFLQELVELKMLAEE----- 697  
QY 204 EGKLIPELNNRLKTQVNFPTSLSATVQKANKDIDAAL--KLATEIAAIGEI---KTET 257  
Db 698 DKRKVSELTSLKLTDEEFQSLKSSHEKSNKSLDKSLFKKSEELAIQLDICKCKTE- 756  
QY 258 ETRFYVDYDMLSLKGAAKMIN 283  
Db 757 -----ALLEAKTNELIN 768  
RESULT 18  
ABUS3077  
ID ABUS3077 standard; protein; 961 AA.  
XX  
AC ABUS3077;  
XX  
DT 15-APR-2003 (first entry)  
XX  
DE Intracellular trafficking-associated DKFZphtes3\_lg13 homologue #10.  
XX  
KW Human; gene therapy; vaccine; disease treatment; detection.  
XX  
OS Homo sapiens.  
XX  
PN WO200112659-A2.  
XX  
PD 22-FEB-2001.  
XX  
PF 18-AUG-2000; 2000WO-IB001496.  
XX  
PR 18-AUG-1999; 99US-0149499P.  
PR 28-SEP-1999; 99US-0156503P.  
XX  
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
PA  
XX  
XX Wiemann S;  
XX  
XX WPI; 2001-327840/34.  
XX  
XX Nucleic acids having the sequences of clones isolated from libraries of  
PT different human tissues, useful in recombinant DNA methodologies.  
PT  
XX  
XX Example III; Page 673; 1095pp; English.  
XX  
XX This invention describes novel polynucleotides and polypeptides isolated  
CC from human cDNA libraries which can be used for gene therapy or in  
CC vaccines. The polynucleotides of the invention and antibodies encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC

CC associated with inappropriate polypeptide expression. The products of the  
CC invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of  
CC the invention may also be used as diagnostic agents for detecting the  
CC presence of polypeptides in samples. This sequence represents a homologue  
CC of a polypeptide described in the disclosure of the invention  
XX  
SQ Sequence 961 AA;  
Query Match 7.3%; Score 110; DB 4; Length 961;  
Best Local Similarity 20.6%; Pred. No. 5.3;  
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;  
QY 7 EOTVEVVKSAIETADGALDLYNKLDQVTPW-KTFDETIKELSRFKQY-----SQEA 58  
Db 297 EKKMEVKQAKEMQETLK--KLLDQEAFLKKELENTALELSQKEQFNKMLEMAQAN 354  
QY 59 SVLVGDIKVLMDSDQKYFRATQVY--EMCGVVT-----QLLSAYILLDFEYN 105  
Db 355 SAGISDAVSRLTQKEQIESLTVHRRELNDVISWEKLNQQAELQEIHIQLOEKE 414  
QY 106 EKKASAQKDILI-----RILDDGVKK--LNEAQSKLLTSSQSFNNASGKLL 149  
Db 415 QEVAELKQKILLFGCEKEEMNKETWLKEGVKQDITLNEQLQKSAHVNSLAQDET 474  
QY 150 ALDSQL-----TNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPFGLLIISYIAAGVI 203  
Db 475 KLAHLEKLEVDLNLKSLKENTFLQELVELKMLAEE----- 510  
QY 204 EGKLIPELNNRLKTQVNFPTSLSATVQKANKDIDAAL--KLATEIAAIGEI---KTET 257  
Db 511 DKRKVSELTSLKLTDEEFQSLKSSHEKSNKSLDKSLFKKSEELAIQLDICKCKTE- 569  
QY 258 ETRFYVDYDMLSLKGAAKMIN 283  
Db 570 -----ALLEAKTNELIN 581  
RESULT 19  
ABR61599  
ID ABR61599 standard; protein; 2228 AA.  
XX  
AC ABR61599;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Human golgin-245 splice variant 1 polypeptide.  
XX  
KW Golgin-245; neurodegenerative disease; neuroprotective; nootropic;  
KW antiparkinsonian; anticonvulsant; gene therapy; human; variant.  
XX  
OS Homo sapiens.  
XX  
PN WO2003087403-A2.  
XX  
PD 23-OCT-2003.  
XX  
PF 16-APR-2003; 2003WO-EF003958.  
XX  
PR 16-APR-2002; 2002EP-00008553.  
PR 16-APR-2002; 2002US-0372424P.  
XX  
XX (EVOT-) EVOTEC NEUROSCIENCES GMBH.  
PA  
XX  
XX Von Der Kammer H, Pohlner J;  
XX WPI; 2003-845345/78.  
XX N-PSDB; ACF58150.  
XX  
XX Diagnosing or prognosticating a neurodegenerative disease comprises  
PT comparing a level or activity of a transcription or translation product  
PT of golgin-245 gene in a test sample with a reference value of a known  
PT disease or health status.  
PT



|      |   |   |                  |          |  |  |
|------|---|---|------------------|----------|--|--|
|      | 204   | EGLKLIPELNRLNTKVONFFTSLSATYQKANOIDAUKL---           | KLATEIAAIGEI---- | KTET 257 |  |  |
| QY   | :   | :   | :                | :        |  |  |
|      | 1167  | DKRKVSELTSKLKTTDEEFQSLKSSEHKSNSLEFKKLSELAIQDICKKTE- |                  |          |  |  |
| DB   | :   | :   | :                | :        |  |  |
|      | 258   | ETTRFYVDVDDLMSLLKGAAKWGIN 283                       |                  |          |  |  |
| QY   | :   | :   | :                | :        |  |  |
|      | 1226  | -----ALLEAKTNELIN 1237                              |                  |          |  |  |
| DB   | :   | :   | :                | :        |  |  |
| <br> |   |   |                  |          |  |  |
|      | RESULT 21   |   |                  |          |  |  |
|      | ABRG61600   |   |                  |          |  |  |
| ID   | ABRG61600   | standard; protein;                                  | 2230 AA.         |          |  |  |
| XX   | AC  |   |                  |          |  |  |
| XX   | ABRG61600;  |   |                  |          |  |  |
| DT   | 15-JAN-2004   | (first entry)                                       |                  |          |  |  |
| XX   |   |   |                  |          |  |  |
| DE   | Human golgin-245 splice variant 2 polypeptide.                            |   |                  |          |  |  |
| KW   | Golgin-245; neurodegenerative disease; neuroprotective; nootropic;        |   |                  |          |  |  |
| KW   | antiparkinsonian; anticonvulsant; gene therapy; human; variant.           |   |                  |          |  |  |
| OS   | Homo sapiens.   |   |                  |          |  |  |
| PN   | WO2003087403-A2.  |   |                  |          |  |  |
| PD   | 23-OCT-2003.  |   |                  |          |  |  |
| XX   | 16-APR-2003; 2003WO-EP003958.   |   |                  |          |  |  |
| PR   | 16-APR-2002; 2002EP-00008553.   |   |                  |          |  |  |
| DR   | 16-APR-2002; 2002US-0372424P.   |   |                  |          |  |  |
| XX   | (EVOT-) EVOTEC NEUROSCIENCES GMBH.  |   |                  |          |  |  |
| PI   | Von Der Kammer H, Pohlner J;  |   |                  |          |  |  |
| XX   | WPI: 2003-845345/78.  |   |                  |          |  |  |
| DR   | N-PSDB; ACF58151.   |   |                  |          |  |  |
| XX   | Diagnosing or prognosticating a neurodegenerative disease comprises       |   |                  |          |  |  |
| PT   | comparing a level or activity of a transcription or translation product   |   |                  |          |  |  |
| PT   | of golgin-245 gene in a test sample with a reference value of a known     |   |                  |          |  |  |
| PT   | disease or health status.   |   |                  |          |  |  |
| XX   | Claim 18; Fig 7; Opp; English.  |   |                  |          |  |  |
| PS   | The invention relates to diagnosing or prognosticating a                  |   |                  |          |  |  |
| CC   | neurodegenerative disease in a subject, or determining whether a subject  |   |                  |          |  |  |
| CC   | is at increased risk of developing the disease. The method involves       |   |                  |          |  |  |
| CC   | determining a level and/or an activity of a transcription or translation  |   |                  |          |  |  |
| CC   | product of a gene coding for golgin-245 and/or a fragment, derivative or  |   |                  |          |  |  |
| CC   | variant in a sample from the subject and comparing the level and/or the   |   |                  |          |  |  |
| CC   | activity to a reference value. The composition and methods are useful in  |   |                  |          |  |  |
| CC   | diagnosing, preventing and/or treating neurodegenerative diseases, such   |   |                  |          |  |  |
| CC   | as Alzheimer's disease, Parkinson's disease, Huntington's disease or      |   |                  |          |  |  |
| CC   | amyotrophic lateral sclerosis. The recombinant, non-human animal is       |   |                  |          |  |  |
| CC   | useful for screening, testing and validating compounds, agents and        |   |                  |          |  |  |
| CC   | modulators in the development of diagnostics and therapeutics to treat    |   |                  |          |  |  |
| CC   | neurodegenerative diseases, in particular Alzheimer's disease. Specific   |   |                  |          |  |  |
| CC   | antibodies are used for detecting the pathological state of a cell in a   |   |                  |          |  |  |
| CC   | sample from a subject. The present sequence represents a human golgin-245 |   |                  |          |  |  |
| CC   | splice variant 2 polypeptide  |   |                  |          |  |  |
| XX   | Sequence 2230 AA;   |   |                  |          |  |  |
| SQ   | Query Match 7.3%; Score 110; DB 7; Length 2230;                           |   |                  |          |  |  |
|      | Best Local Similarity 20.6%; Pred. No. 16;                                |   |                  |          |  |  |
|      | Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;          |   |                  |          |  |  |
| QY   | 7 EQTVVEVKSALETADGALDLYNKYLDOVIPW-KTFDETIKELSRFQBY-----SQEA 58            |   |                  |          |  |  |
| DB   | 953 EKKGVKKQAKENQMFTLK--KKLLDDQAKLKELENTALESQEKOFNAKMLEMAQAN 101          |   |                  |          |  |  |









Db 425 VTETFRSISGLVEMANDLEKRALAQVSKDYTEAI--NQVNEAIQOVSEIARQOQETI 482  
Qy 59 SVLVGDIKVLMLDSQDKYFEATQTVVEMCGVVTQLLSAYILLDFEYNEKASQKDI--- 115  
Db 483 NEITDGMRLVAQTSE---ESVRAMEEFGAVTEVVS---IANEGSQKGDEALKRIEDI 534  
Qy 116 --LIRILDDGVKVLNEAQKSL-----LTSSQSPNNASGKLLALDS----- 153  
Db 535 QHMSRIEETSVKVAEMSRNIEITNVITSIAQTN-----LLALNAAIEAARAGEAGRGF 590  
Qy 154 -----QUTNDFSEKSSYFQSQVDRIR---KEAVAGAAAGI-VAGPFGLLIISYISIAAG 201  
Db 591 AVVAQEIRKLAESKQAAADNIKSIIDKITDEIKEAVEATKEGVSIGESSETLRDTIGYL 650  
Qy 202 VIGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAUKLATEIAAIGETITETTR 261  
Db 651 ANIATLIQETSERMTTVKE-----QIVRTQBEVDKALRALENLAASAEETTASAEVS 703  
Qy 262 FYVDYDDLMLSLKGAKKMINTCNEYQORHGK 294  
Db 704 SAIEQQTAAIEELRRAAQELKDMVGRMRQIVGK 736

## RESULT 26

ABU53079

ID ABU53079 standard; protein; 929 AA.

XX AC ABU53079;

XX DT 15-APR-2003 (first entry)

XX DE Intracellular trafficking-associated DKFZphtes3\_1913 homologue #12.

XX DE Human; gene therapy; vaccine; disease treatment; detection.

XX DE Homo sapiens.

XX PN W0200112659-A2.

XX PD 22-FEB-2001.

XX PF 18-AUG-2000; 2000WO-IB001496.

XX PR 18-AUG-1999; 99US-0149499P.

XX PR 28-SEP-1999; 99US-0156503P.

XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.

XX PI Wiemann S;

XX DR WPI; 2001-327840/34.

XX DR Nucleic acids having the sequences of clones isolated from libraries of

XX PT different human tissues, useful in recombinant DNA methodologies.

XX PS Example III; Page 674-675; 1095pp; English.

XX CC This invention describes novel polynucleotides and polypeptides isolated

XX CC from human cDNA libraries which can be used for gene therapy or in

XX CC vaccines. The polynucleotides of the invention and antibodies encoded by

XX CC them may be used in the prevention, diagnosis and treatment of diseases

XX CC associated with inappropriate polypeptide expression. The products of the

XX CC invention may also be used to identify modulators of expression and

XX CC activity and to down regulate expression and activity. The antibodies of

XX CC the invention may also be used as diagnostic agents for detecting the

XX CC presence of polypeptides in samples. This sequence represents a homologue

XX CC of a polypeptide described in the disclosure of the invention

XX SQ Sequence 929 AA;

XX SQ Query Match 7.28; Score 109; DB 4; Length 929;

XX SQ Best Local Similarity 20.6%; Pred. No. 6.2;

Matches 67; Conservative 59; Mismatches 106; Indels 94; Gaps 12;  
Qy 7 EQTVEVVKSAIETADGALDLYNKYLQVTPW-KTFDETIKELSRFKOEY-----SOBA 58  
Db 1 EKVQKAKEMOET-----LKKLLDQEAQKKELENTALELSQKEQFNAKMLEMAQAN 54  
Qy 59 SVLVGDIKVLMLDSQDKYFEATQTVV--EWCQVVT-----QLLSAYILLDFEYN 105  
Db 55 SAGISDAVSRLETNQKEQIESLTVHRRELNDVISIWEKKLNOQAELQIHEIQLOEKE 114  
Qy 106 EKKASQAQDILI-----RILDDGVK--LNEAQSLTSSQSPNNASGKLL 149  
Db 115 QEVAELQKQILLPGCEKEEMNKETIWLKEBQVQDTTLNELQBLQKQSAHVNSLAQDET 174  
Qy 150 ALDSQL-----TNDSEKSSYFQSQVDRIRKEAYAGAAAGIVAGPFGLLIISIAAGVI 203  
Db 175 KKAHLKLEKLEVDLNLKSLKENTFLQEQVLKMLAEE----- 210  
Qy 204 EGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAUKL-----KLATEIAAIGEI---KTET 257  
Db 211 DRKVSLETSKLTDTDEBFQSLKSSHEKSNKSLSDKSLFKLSELAIQDIDICCKYTE- 269  
Qy 258 ETTRFYVDYDDLMLSLKGAKKMIN 283  
Db 270 -----ALLEAKTNELIN 281

## RESULT 27

ABB59245

ID ABB59245 standard; protein; 1048 AA.

XX AC ABB59245;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 4527.

XX DE Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN W0200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL03348.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX PT interactions.

XX PS Disclosure; SEQ ID NO 4527; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent

XX CC capable of detecting 1000 or more genes from Drosophila. The invention is

XX CC useful in developmental biology and in elucidating cell signalling and

XX CC cell-cell interactions in higher eukaryotes for the development of

XX CC insecticides, therapeutics and pharmaceutical drugs. The invention

XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

XX CC ABB72072). The sequence data for this patent did not form part of the

XX CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1048 AA;

Query Match 7.2%; Score 108.5; DB 4; Length 1048;

Best Local Similarity 21.6%; Pred. No. 7.9; Mismatches 110; Indels 53; Gaps 12;

Matches 62; Conservative

QY 24 LDLYNKYLDQVLPKTKDEIKELSRFKQYSEASVLDGDIKVLMDSDQKYFEATQTV 83

DB 159 LELYNELCDLL---STDDTK--IRIFDSTKKGSVIIQGLEIPVHSDVDVYKLEKG 213

QY 84 YEWGCVVTLQLSAYILLFDEYNEKASAKDILIRILDGVK-----KLENAQKSLTSS 138

DB 214 KERRKTATTLNNA-----QSSRSHTVFSIVVHIRENGIEGEDMLKIGKLNLDLAGS 265

QY 139 QSFNNASG-----LLALDSOLTN--DFSEKSSYFQSOVDRIKRAYAGAA 182

DB 266 ENVKAGNEKIGIRVRETWNINQSLTLGRVITLVDRAPHVPYRESKLTLLQESLGRT 325

QY 183 -----AGIVAGPFGI-----IISYSIAAGVIEGKLIPELNNRL--KTVQNFPTLSATVK 230

DB 326 KTSIIATISPGHKDIEETLSLEYAHRANKIQNK--PEVNQKLTKTVLKEYTE---EID 380

QY 231 QANKDIDAUKLATEIA--AIGEIKTETETTRFVVDYDDMLSLK 275

DB 381 KLRDLMAARDKNGIYLAETTYGEITLKESQNLRELNEKMLLLKALK 427

RESULT 28

AAG67418

ID AAG67418 standard; protein; 1066 AA.

XX AC AAG67418;

DT 13-NOV-2001 (first entry)

DE Amino acid sequence of bimC homologue, KLP61.

XX bimC; kinesin related protein; fungal viability; antifungal; KLP61;

KW fungal infection.

XX Drosophila melanogaster.

XX US6284480-B1.

XX 04-SEP-2001.

XX 03-APR-2000; 2000US-00541782.

XX 03-APR-2000; 2000US-00541782.

XX (CYTO-) CYTOKINETICS INC.

XX Nislow CE, Sakowicz R, Beraud C;

XX WPI; 2001-540724/60.

XX N-PSDB; AAH78013.

XX Identifying a modulator, e.g. antifungal agent, of a target protein

PT comprising bimC or its fragment by determining enzymatic activity of a

PT reaction, in the presence and absence of the compound, that uses ADP or

PT phosphate produced by bimC.

XX Disclosure; Fig 8; 47pp; English.

XX The present sequence represents a bimC homologue, designated KLP61. BimC

CC is a kinesin related protein, which is essential for fungal viability.

CC The specification describes a method of identifying modulators of bimC.

CC The method comprises adding a test agent to a mixture comprising bimC

CC protein that directly or indirectly produces ADP or phosphate, subjecting

CC the mixture to an enzymatic reaction that uses the ADP or phosphate, and

CC determining the enzymatic activity in presence and absence of test

CC compound. A change in the activity level between the presence and absence

CC of the candidate agent indicates a modulator of the target protein

CC function. The method is useful for identifying a modulator, e.g.

CC antifungal agents, of bimC. The modulators can be used, for example, to

CC inhibit the growth or spread of fungi, mould, fruit flies, etc. . The

CC modulators can be used for preventing and treating infections caused by

CC Chytridiomycetes, Hyphochytridiomycetes, Plasmodiophoromycetes,

CC Oomycetes, Zygomycetes, Ascomycetes, and Basidiomycetes

XX SQ Sequence 1066 AA;

Query Match 7.2%; Score 108.5; DB 4; Length 1066;

Best Local Similarity 21.6%; Pred. No. 8.1;

Matches 62; Conservative

QY 24 LDLYNKYLDQVLPKTKDEIKELSRFKQYSEASVLDGDIKVLMDSDQKYFEATQTV 83

DB 159 LELYNELCDLL---STDDTK--IRIFDSTKKGSVIIQGLEIPVHSDVDVYKLEKG 213

QY 84 YEWGCVVTLQLSAYILLFDEYNEKASAKDILIRILDGVK-----KLENAQKSLTSS 138

DB 214 KERRKTATTLNNA-----QSSRSHTVFSIVVHIRENGIEGEDMLKIGKLNLDLAGS 265

QY 139 QSFNNASG-----LLALDSOLTN--DFSEKSSYFQSOVDRIKRAYAGAA 182

DB 266 ENVKAGNEKIGIRVRETWNINQSLTLGRVITLVDRAPHVPYRESKLTLLQESLGRT 325

QY 183 -----AGIVAGPFGI-----IISYSIAAGVIEGKLIPELNNRL--KTVQNFPTLSATVK 230

DB 326 KTSIIATISPGHKDIEETLSLEYAHRANKIQNK--PEVNQKLTKTVLKEYTE---EID 380

QY 231 QANKDIDAUKLATEIA--AIGEIKTETETTRFVVDYDDMLSLK 275

DB 381 KLRDLMAARDKNGIYLAETTYGEITLKESQNLRELNEKMLLLKALK 427

RESULT 29

ABG72693

ID ABG72693 standard; protein; 1066 AA.

XX AC ABG72693;

DT 06-MAR-2003 (first entry)

DE Fruitfly KLP61, a homologue of BimC, a kinesin related protein.

XX Fruitfly; bimC; kinesin related protein; mitosis; antifungal;

KW Candida infection; fusaria infection; zygomycosis; cryptococcosis;

KW histoplasmosis; KLP61.

XX Drosophila melanogaster.

XX US6468760-B1.

XX 22-OCT-2002.

XX 28-NOV-2000; 2000US-00723820.

XX 03-APR-2000; 2000US-00541782.

XX (CYTO-) CYTOKINETICS INC.

XX Nislow CE, Sakowicz R, Beraud C;

XX WPI; 2003-147102/14.

XX N-PSDB; ABX14663.

XX Novel kit for screening for modulators of bimC, comprises a biologically

PT active bimC, its fragment or homolog, and instructions for testing bimC

PT activity.

XX Disclosure; Fig 8; 47pp; English.



PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA47181.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 71235; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 3692 AA;  
Query Match 7.2%; Score 108.5; DB 6; Length 3692;  
Best Local Similarity 21.5%; Pred. No. 40;  
Matches 71; Conservative 49; Mismatches 128; Indels 83; Gaps 12;  
QY 11 EVVKSATETADGALDLYNKYLQVVPKTFDETIKLSR-----FKQEQSQ-EA 58  
DB 564 EQVKNDIIPSNYTLASYNKY-----NKLKRAQTVDLTETNTFPNQRYSQTOI 612  
QY 59 SVLVGDIKVLMM-----DSQKYPEATQTVYEWGVVTTQLLSAYILLFDEYNEKKASA 111  
DB 613 DDLHLQLTTLINRVASREINDKAQEMTDAVYDSTELTTEKDT---LVDQIENHKGNEI 669  
QY 112 QKDILRIILDGKVKLNEAKSLTSS-----QSFNNSAGKLLALDSQLTNDFS 160  
DB 670 SNNDDELTDGQVVERKVEAGLHTLES DTPHVTKPNARQVNNRA-----DQCKTLIRN 723  
QY 161 EKSSYFOSQVDRIRK-EAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTQV 219  
DB 724 NHEATTEEQNEAIRQVEAHSSDAIAKIGE-----AETDTTVNEAR 763  
QY 220 NFFTSLSAT-VKQANKIDIAKLAETAAIGETETETTFYVDYDMLSLKGA 278  
DB 764 DNGTKLIATDVPNPTKKAEEA---RAAVTNSANSKIDINNNTQTDLERDAIALVNRSK 820  
QY 279 KQIMTNCNEYQ-----QRHGKTKLFEVP 301  
DB 821 DEALQNTAQGNDVDVTEAQNGNTTIQQVP 851

RESULT 32  
ABP40235  
ID ABP40235 standard; protein; 3696 AA.  
XX  
XX AC ABP40235;  
XX DT 24-JUL-2002 (first entry)  
XX  
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5080.  
DE  
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy.  
KW Staphylococcus epidermidis.  
OS  
XX US6380370-B1.  
XX  
XX 30-APR-2002.  
XX  
XX 13-AUG-1998; 98US-00134001.  
XX  
XX 14-AUG-1997; 97US-0055779P.  
PR 08-NOV-1997; 97US-0064964P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Doucette-Stamm LA, Bush D;  
PI WPI; 2002-381255/41.  
DR N-PSDB; ABN92780.  
XX  
XX Novel isolated nucleic acid encoding a Staphylococcus epidermis  
PT polypeptide, useful for diagnosing and treating bacterial infections.  
XX  
XX Disclosure; SEQ ID NO 5080; 267pp; English.  
XX  
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences can  
CC also be used in the diagnosis and treatment of bacterial infections,  
CC particularly *S. epidermidis* infections. The sequences can be used to  
CC screen for compounds able to interfere with the *S. epidermidis* life cycle  
CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from the USPTO web site  
XX SQ Sequence 3696 AA;

Query Match 7.2%; Score 108.5; DB 5; Length 3696;  
Best Local Similarity 21.5%; Pred. No. 40;  
Matches 71; Conservative 49; Mismatches 128; Indels 83; Gaps 12;  
QY 11 EVVKSATETADGALDLYNKYLQVVPKTFDETIKLSR-----FKQEQSQ-EA 58  
DB 568 EQVKNDIIPSNYTLASYNKY-----NKLKRAQTVDLTETNTFPNQRYSQTOI 616  
QY 59 SVLVGDIKVLMM-----DSQKYPEATQTVYEWGVVTTQLLSAYILLFDEYNEKKASA 111  
DB 617 DDLHLQLTTLINRVASREINDKAQEMTDAVYDSTELTTEKDT---LVDQIENHKGNEI 673  
QY 112 QKDILRIILDGKVKLNEAKSLTSS-----QSFNNSAGKLLALDSQLTNDFS 160  
DB 674 SNNDDELTDGQVVERKVEAGLHTLES DTPHVTKPNARQVNNRA-----DQCKTLIRN 727  
QY 161 EKSSYFOSQVDRIRK-EAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTQV 219  
DB 728 NHEATTEEQNEAIRQVEAHSSDAIAKIGE-----AETDTTVNEAR 767  
QY 220 NFFTSLSAT-VKQANKIDIAKLAETAAIGETETETTFYVDYDMLSLKGA 278

Db 768 DNGTKLIATVPNPTKAEAA---RAAVTNSANSKIKDINNNTQATILDERNDATAVNRSK 824

Qy 279 KMINTCNEYQ-----QRHGKTLFEVP 301

Db 825 DEAIQININTAGNDVTEAQNNGTNTIQQVP 855

RESULT 33

ABB77986

XX ID ABB77986 standard; protein; 1292 AA.

XX AC ABB77986;

XX XX

DT 22-OCT-2002 (first entry)

XX XX

DE Amino acid sequence of an Arabidopsis RAD50 homologue.

XX Nucleic acid integration; homologous recombination; telomeric region;

KW RAD50.

XX Arabidopsis thaliana.

OS

XX

PN EP1217074-A1.

XX

PD 26-JUN-2002.

XX

PF 22-DEC-2000; 2000EP-00204693.

XX

PR 22-DEC-2000; 2000EP-00204693.

XX

XX (UYLE-) RIJKSUNIV LEIDEN.

PA (BINA-) STICHTING BINAIR VECTOR SYSTEM.

XX

PI Hooykaas PJJ, Van Attikum H, Bundock P;

XX WPI; 2002-550405/59.

DR

XX

PT Directing integration of nucleic acid of interest to a sub-telomeric

PT region in an eukaryote with preference for non-homologous recombination,

PT by steering an integration pathway towards homologous recombination.

XX

PS Disclosure; Fig 5; 63pp; English.

XX

CC The specification describes a method for directing integration of a

CC nucleic acid of interest to a pre-determined site, where the nucleic acid

CC has homology at or around the pre-determined site, in a eukaryote with a

CC preference for non-homologous recombination. The method comprises

CC steering an integration pathway towards homologous recombination. The

CC method is useful for directing integration of a nucleic acid of interest

CC to a subtelomeric and/or telomeric region in an eukaryote with a

CC preference for non-homologous recombination. The nucleic acid of interest

CC comprises an inactive gene to replace an active gene, or vice versa, is a

CC portion of a gene delivery vehicle, confers a desired property to the

CC eukaryotic cell, or encodes a therapeutic proteinaceous substance or a

CC substance conferring resistance for an antibiotic substance to a cell.

CC The method is useful for improving gene targeting efficiency. The method

CC is useful in the replacement of an active gene by an inactive gene, for

CC e.g. for the inactivation of genes controlling undesired side branches of

CC metabolic pathways, to increase the quality of bulk products such as

CC starch, or to increase the production of specific secondary metabolites

CC or to inhibit formation of unwanted metabolites, and also to inactivate

CC genes controlling senescence in fruits and flowers or that determine

CC flower pigments. The method is also useful for replacing an inactive gene

CC by an active gene. For e.g. the replacement of a defective p53 by an

CC intact p53. Many tumours acquire a mutation in p53 during their

CC development which renders it inactive and often correlates with a poor

CC response to cancer therapy. By replacing the defect p53 by an intact p53,

CC e.g. through gene therapy, conventional anti cancer therapy have better

CC changes of succeeding. The method is also useful for therapeutic

CC proteinaceous substance integration. A tumoricidal gene can be delivered

CC to a pre-determined site present only in e.g. proliferating cells, or

CC present only in tumour cells, e.g. to the site where a tumour antigen is

CC expressed form. ABB77984-86 represent RAD50 homologues. RAD50 is involved

CC

CC in non-homologous recombination

XX

SQ Sequence 1292 AA;

Query Match 7.1%; Score 107.5; DB 5; Length 1292;

Best Local Similarity 17.2%; Pred. No. 13;

Matches 69; Conservative 77; Mismatches 138; Indels 117; Gaps 13;

Qy 2 TSIFABQTVVEVKSATETADGALDLYNKYLDQVIPWKT-PDE----- 42

Db 281 STLFKEQOQYAAALPEENED-----TTEELKEWKSFEERLALGTKIRKMEREMVD 332

Qy 43 ---TIKELSRFKQYEQEASVL---VGDIKVLLWDSQDKYFEATQTVYEWGVVTTQLLSAY 97

Db 333 TETTISLHNKATNYMLEISKLQTEAEAHMLLNKRDSTIQTFFHNLGNVSTPSTPSTE 392

Qy 98 ILLFDEYNEKKASQAQDILIRILDGKVKLNEAQSLLTSSQSFNNASGKLLALDSQ--- 154

Db 393 VVL--NLTNRIKSRGLGELEMDLLDK--KKSNET--ALSTAWDCYMDANDRWKSEIAQKBA 446

Qy 155 -----LTNDFSEKSSY-----FQSQ 169

Db 447 KDEIKMGISKRIEKEIERDSFEFEISTVDVKQTDEREKQVQVELEKTKQNSRGEFSK 506

Qy 170 VDIRIRKEAYA-----GAAAGIVAGPFGLLIISYSI-----AAGVIEGKLIPELNNR- 214

Db 507 IEQKHIEIVSLEHKIKTLNRERDVMDAGDAEDRLLTRIDECKDIRGVGLKRLPEKDKMR 566

Qy 215 -----LKTQNFPTSLSATVKQANKDIDAALKLATEIAAIGEIKTETTRFFVVDYD-- 267

Db 567 EIVQALRSIEREYDDLSLSREAKEVNMLOMKIQEVNNSLFPHNKDTERSKRYIESKLQ 626

Qy 268 -----DLMLSLKGAAGKMMINTCNEYQQRHGKTLFE 299

Db 627 ALKQESVTIDAYPKLLESACKDKDRKREYNMANGRMQMEF 667

RESULT 34

ABR53560

ID ABR53560 standard; protein; 1875 AA.

XX AC ABR53560;

XX XX

DT 20-JUN-2003 (first entry)

XX XX

DE Protein sequence #SEQ ID 1985.

XX XX

KW Multiprotein complex; eukaryote; drug target; diagnosis.

XX XX

OS Saccharomycetes cerevisiae.

XX XX

PN EP1258494-A1.

XX XX

PD 20-NOV-2002.

XX XX

PF 20-DEC-2001; 2001EP-00130253.

XX XX

PR 15-MAY-2001; 2001EP-00111774.

XX XX

PA (CELL-) CELLZOME AG.

XX XX

PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;

PI Marzioch M, Schultz JD, Superti-Furga GD;

XX XX

DR WPI; 2003-250078/25.

DR N-PSDB; ACC61602.

XX XX

PT New isolated protein complexes useful for diagnosing a disease or

PT disorder, or as a target for an active agent of a pharmaceutical,

PT preferably a drug target in the treatment or prevention of disease or

PT disorder.

XX

PS Disclosure; SEQ ID NO 1985; 17pp + Sequence Listing; English.

XX The invention relates to multiprotein complexes from eukaryotes. Proteins  
CC of the invention and DNA sequences encoding them are given in records  
CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are  
CC obtainable by using a protein as a bait and isolating the set of proteins  
CC which is attached thereto from cells. Such protein complexes may comprise  
CC up to 30 distinct proteins. Protein complexes of the invention are useful  
CC for diagnosing a disease or disorder, or as a target for an active agent  
CC of a pharmaceutical, preferably a drug target in the treatment or  
CC prevention of a disease or disorder. Note: The sequence data for this  
CC patent is not represented in the printed specification, but is based on  
CC sequence information supplied by the European Patent Office. The complete  
CC document is available on CD-ROM  
XX  
XX Sequence 1875 AA;  
Query Match 7.1%; Score 107.5; DB 6; Length 1875;  
Best Local Similarity 19.9%; Pred. No. 20;  
Matches 64; Conservative 49; Mismatches 127; Indels 81; Gaps 12;  
QY 7 EQTVVVKSAIETADGALDLYNKYLDQVVPWKTFDETI-----KELSRFKOEYSQEAS 59  
Db 966 EDKISLLKEQMFNLNNELDLQKGMKEK--EKADFKKRISILQNNKNEVEAVKSEYSEKLS 1023  
QY 60 VLVDGIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKASAKDI---- 115  
Db 1024 KIQND-----LDQOTIYANTQNNYE-----QELQKHADVSKTISELR 1061  
QY 116 -LIRILDGVKKLN-----EAQKSLTTSQSFNNASGKLLA---LDSQLTNDPSEKSYFQ 167  
Db 1062 EQHTYKGVQVKTNLNLRDQLENALKENEKSWSSQKESLLEQLDLSNRIEDLSQNKLLY 1121  
QY 168 SQVDRIRKEAYAGAAGIVAGPFGGLIISYSTAAGVIEGKLIPELNNRLKTVQNFFTSLSA 227  
Db 1122 DQI-----QIYTAADKEVNNSTNG-----PGLNNILITLRERDILT 1159  
QY 228 TVQKANKDIDAAKKLATEIAAIGETKTETETTR-----FVVDYDDL-----LSLL 274  
Db 1160 KTVVAERDAKMLRQKISLMDVELQDARTKLDNSRVEKENHSSIIQQHDDIMEKLNQLNLL 1219  
QY 275 KGAAKKMINTCNEYQQRHGKK 295  
Db 1220 R---ESNITLRNELENNNKK 1237  
RESULT 35  
ADK64380  
ID ADK64380 standard; protein; 1875 AA.  
XX  
AC ADK64380;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Disease treating protein complex-derived protein #1185.  
XX  
KW protein complex; drug target; diagnosis.  
XX  
OS Unidentified.  
XX  
XX EP1338608-A2.  
XX  
XX 27-AUG-2003.  
XX  
XX 20-DEC-2002; 2002EP-00102902.  
XX  
XX 20-DEC-2001; 2001EP-00130253.  
XX  
XX (CELL-) CELLZOME AG.  
XX  
XX Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;  
XX Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;  
XX Michon A, Leutwein C, Rick J;  
XX

DR WPI; 2003-638460/61.  
XX N-PSDB; ADK64381.  
PT New proteins and protein complexes from eukaryotes, useful as targets in  
PT drug screening, or in diagnosing or screening for the presence of a  
PT disease or disorder, or a predisposition for developing a disease or  
PT disorder in a subject.  
XX  
PS Disclosure; SEQ ID NO 2369; 13pp; English.  
XX  
CC The invention relates to novel protein complexes comprising a first and a  
CC second protein, or its derivative, fragment, homologue or variant. The  
CC proteins are selected from given protein complexes, which are not defined  
CC in the specification. The variants are encoded by nucleic acids that  
CC hybridize to the nucleic acids encoding the proteins under low stringency  
CC conditions. The protein complexes are useful as targets for an active  
CC agent of a pharmaceutical. These protein complexes are particularly  
CC useful as drugs targets for the treatment or preventing of a disease or  
CC disorder. The complexes and methods above are useful in diagnosing or  
CC screening for the presence of a disease or disorder. These are also useful  
CC for developing a drug for treatment or prevention of a disease or  
CC in screening for a drug for treatment or prevention of a disease or  
CC disorder. The molecule that modulates the amount, activity or protein  
CC components of the complex is useful for the manufacture of a medicament  
CC for the treatment or prevention of a disease or disorder. This sequence  
CC corresponds to a protein of the invention. (Note: the sequence data for  
CC this patent did not form part of the printed specification but was  
CC obtained from the EPO in electronic format).  
XX  
SQ Sequence 1875 AA;  
Query Match 7.1%; Score 107.5; DB 7; Length 1875;  
Best Local Similarity 19.9%; Pred. No. 20;  
Matches 64; Conservative 49; Mismatches 127; Indels 81; Gaps 12;  
QY 7 EQTVVVKSAIETADGALDLYNKYLDQVVPWKTFDETI-----KELSRFKOEYSQEAS 59  
Db 966 EDKISLLKEQMFNLNNELDLQKGMKEK--EKADFKKRISILQNNKNEVEAVKSEYSEKLS 1023  
QY 60 VLVDGIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKASAKDI---- 115  
Db 1024 KIQND-----LDQOTIYANTQNNYE-----QELQKHADVSKTISELR 1061  
QY 116 -LIRILDGVKKLN-----EAQKSLTTSQSFNNASGKLLA---LDSQLTNDPSEKSYFQ 167  
Db 1062 EQHTYKGVQVKTNLNLRDQLENALKENEKSWSSQKESLLEQLDLSNRIEDLSQNKLLY 1121  
QY 168 SQVDRIRKEAYAGAAGIVAGPFGGLIISYSTAAGVIEGKLIPELNNRLKTVQNFFTSLSA 227  
Db 1122 DQI-----QIYTAADKEVNNSTNG-----PGLNNILITLRERDILT 1159  
QY 228 TVQKANKDIDAAKKLATEIAAIGETKTETETTR-----FVVDYDDL-----LSLL 274  
Db 1160 KTVVAERDAKMLRQKISLMDVELQDARTKLDNSRVEKENHSSIIQQHDDIMEKLNQLNLL 1219  
QY 275 KGAAKKMINTCNEYQQRHGKK 295  
Db 1220 R---ESNITLRNELENNNKK 1237  
RESULT 36  
ABP39975  
ID ABP39975 standard; protein; 1211 AA.  
XX  
AC ABP39975;  
XX  
XX 24-JUL-2002 (first entry)  
XX  
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4820.  
XX  
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.  
XX

```

OS Staphylococcus epidermidis.
XX US6380370-B1.
XX 30-APR-2002.
XX 13-AUG-1998; 98US-00134001.
XX 14-AUG-1997; 97US-0055779P.
XX 08-NOV-1997; 97US-0064964P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm LA, Bush D;
XX WPI; 2002-381255/41.
XX N-PSDB; ABN92520.
XX Novel isolated nucleic acid encoding a Staphylococcus epidermis
XX polypeptide, useful for diagnosing and treating bacterial infections.
XX Disclosure; SEQ ID NO 4820; 267pp; English.
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences can
XX also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life cycle
XX or inhibit S. epidermidis infection. N.B. The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from the USPTO web site
XX
SQ Sequence 1211 AA;
Query Match 7.1%; Score 107; DB 5; Length 1211;
Best Local Similarity 18.3%; Pred. No. 13;
Matches 61; Conservative 59; Mismatches 131; Indels 82; Gaps 9;

QY 7 EQTVVVK-----SAIETADGALDLYNKYLDQVLPWKTFDETIKLSRFKQYSEASV 60
D 185 EESAGVLKYKKRAESIQKLDHTEDNLNRVEDILYDLEGRVPLKEEAATAKEYKQLSKE 244
QY 61 LVGDIKVLMDSQKYPEATQTVVWC-----GVVTQLLSAYILLFDYNE 106
D 245 MEQSDVITVTSIDHYTEDNORLDERLNHLKSQAQKEGQAQINQLLQKY----- 295
QY 107 KKASQAKDILIRILDGKVLNEAKSLTSSQSFNNASGKLLALDSQLTND-----FS 160
D 296 -KGRQON-----DYDIEKLN-----YELVKATENYEQLSKLVLEERKKNQSETNARYE 345
QY 161 EKSSYFQSDVRIRKEAYAGAAAGIVAGPFGLIITSYIAAGVTEGKLIPELNRL-----K 216
D 346 BELDNLSEQIDSINKKAQNE-----KLLADLKNKQKQLNK 381
QY 217 TVQNFFTSLSATVKQANKIDAAKKLATIAAIGEIKTETTRFYVDVDDLMLSLKG 276
D 382 EVQELSLLYISDQHEKLEEKNSYITLMSQSVVNNDRIFLEHTINENEAKSRDLS 441
QY 277 AAKQWINTCNEYQ-----RHGKTLFEV 300
D 442 RLVEAFNQLKDIOQNTITQTOKEYOSSKSMKV 474

RESULT 37
ABB49720
ID ABB49720 standard; protein; 927 AA.
XX AC ABB49720;
XX DT 05-FEB-2002 (first entry)
XX

```

---

```

DE Listeria monocytogenes protein #2424.
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX vitamin B12; bacterial infection; disease.
XX Listeria monocytogenes.
XX WO200177335-A2.
XX 18-OCT-2001.
XX 11-APR-2001; 2001WO-FR001118.
XX 11-APR-2000; 2000FR-00004629.
XX (INSP ) INST PASTEUR.
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
XX Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
XX Daniels J, Gobbel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA, A;
XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend L;
XX Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
XX Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
XX Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
XX Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and related
XX polypeptides.
XX Claim 6; SEQ ID NO 2425; 192pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
XX monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
XX it are useful for selecting probes and primers for detecting genes in L.
XX monocytogenes and related organisms, and for studying genetic
XX polymorphisms and other genomes. The present sequence is a protein
XX encoded by the genome sequence of the present invention. Proteins
XX expressed from the genome sequence are useful for raising specific
XX antibodies, identification of L. monocytogenes and related organisms, and
XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin
XX B12. The genome sequence and proteins encoded by it are also useful for
XX selecting compounds that regulate gene expression and cell replication
XX and modulate L. monocytogenes-related diseases. In addition, the genome
XX sequence and proteins encoded by it are useful in pharmaceutical and
XX vaccine compositions for the treatment or prevention of infections by L.
XX monocytogenes and related organisms. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 927 AA;
Query Match 7.0%; Score 106.5; DB 5; Length 927;
Best Local Similarity 20.5%; Pred. No. 9.8;
Matches 78; Conservative 44; Mismatches 131; Indels 127; Gaps 16;

QY 1 MTSIFAEQTVVVKSAIETADG---ALDLYNKYLDQVLPW---KTFDETIKLSRFKQE 53
D 160 VTKSYAEAFDKIK---ESGDGFAQADGSGKIKDGLVKQEGNKTISTNLKTLADSSLT 216
QY 54 YSQEASVL-----VGDIK-----VLLMDS 72
D 217 FKDGANTLEVGLTYTDGVNNTAAAGDKLNAGVSTLAAGVPLKDGVAALDGGATKLASG 276
QY 73 QDKYFATQTVVYEWCGVVTQLLSAYILLFDYNEKKASQAKDILIRILDGKVLNEAK 132
D 277 VSTYTSQVDTL---AGGINQAYTGSTALSDGLNRMNGS-----VPTLASGITQLNNGQK 327
QY 133 SLITSSQSFNNASGKLLA---LDSQLTN-----DFSEKSSYFQSDVRIRKEAYAGAA 182

```

328 SLATGLDSLVDGSKLSAGLKELDGNLTKQKIAQLKQGMNDLQGLDQLNQSVNGEDA 387  
 183 AGIVAGPFGLLIISYSTAAGVIEGKLIPELNRL-----KTQVNFFTSLSAT 228  
 388 A-----LAKLAT-----LQKSLDILQGLTFIKSNANFDAEAIKSKINATAGVSAE 434  
 229 VKQ-----ANKDIDAAKLKLATIAAIGIKT-----ETETTRFYVDYDMLLSL 273  
 435 DKQIIDAQADLDKETQKSAQTQVATVEQLQSLGLDLAAITQV-----EL 483  
 274 LKGAAKKQKMTNCEYQORHG 293  
 484 QTGVAKISAG-----YQAVHG 499

RESULT 38  
 AAW22775  
 ID AAW22775 standard; protein; 1312 AA.  
 XX AC AAW22775;  
 XX DT 21-DEC-1998 (first entry)  
 XX DE Human RAD50.  
 XX KW Human; RAD50; DNA repair; tumour suppression; cancer; Septin-2;  
 KW central nervous system.  
 XX OS Homo sapiens.  
 XX PN WO9727284-A2.  
 XX PD 31-JUL-1997.  
 XX PF 24-JAN-1997; 97WO-US001299.  
 XX PR 26-JAN-1996; 96US-00592126.  
 PR 17-JUL-1996; 96US-00687080.  
 XX PA (GENE-) GENELABS TECHNOLOGIES INC.  
 XX PI Dolganov G;  
 XX PS WPI; 1997-393672/36.  
 DR N-PSDB; AAT75237.  
 XX PT Human tumour suppressor gene RAD50 - useful to detect predisposition to,  
 PT decrease risk of and treat cancer, also Septin-2 homologues.  
 XX PS Claim 5; Page 82-86; 195pp; English.  
 XX CC The human RAD50 (hRAD50) is involved in DNA repair and has tumour  
 CC suppression activity, can be used to detect predisposition to, decrease  
 CC the risk of or treat cancers, e.g. acute myeloid leukaemia,  
 CC myelodysplastic syndrome, therapy related myelodysplastic syndrome,  
 CC therapy related acute myeloid leukaemia, refractory anaemia or refractory  
 CC anaemia with excess blasts. Also disclosed in this invention is Human  
 CC Septin-2 homologues of which may be used as targets for cancer therapies  
 CC and central nervous system directed treatment methods, and to measure the  
 CC proliferative potential of selected cell types  
 XX SQ Sequence 1312 AA;  
 Query Match 7.0%; Score 106.5; DB 2; Length 1312;  
 Best Local Similarity 18.9%; Pred. No. 15;  
 Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;  
 QY 10 VEVKSAJETADGALDLYNKYLDQVLPWKFDETIK-----ELSRFKQY 54  
 502 MEVISLQNEKAD--LDRLRLKLDQEMELNHHHTTQMEMLTKQKADKDEQIRKISRH 559  
 QY 55 SOEASVLGDIKVLMLDSQDKYFEATQVYEWCGVVTQLLSAYILLFDEYNEKKASAKD 114

560 SDELTSLG-----YFPNKKQLEDWLHKSXK-----EINQTR----- 591  
 115 ILRLDLDGKVLNEAQKILLTSSQSNFNASGKLALDSQLTNDSEK-----SSYPOS 168  
 592 -----DRAKUN-----KELASSEQNKHINNELKKEQLSS-YEDKLFDVCGSQDFES 640  
 169 QVDRIKRE-----AVAGAAAGIVAGPF----- 190  
 641 DLDELKEEIEKSKQRAMLAGATA--VYSQFITQLTDENOSCCPVCQRFQTEAELEQVI 698  
 191 -----GLI-ISYSIAAGVIEGKLIPELNRLKT 217  
 699 SDLOSRLRLAPDKLSTESLKKKKERDEMGLVPMRQSIID--LKEKEIPELRNKLQV 756  
 218 VQNFETSLSATVQANKDIDAAKLKLATIAAIGIKTETET-----RVVDY 266  
 757 V-----NRDIQRLKNDIBEQETLLGTIMPEESAKVCLTDVTIMERFQML 802  
 267 DDLMLSLKGAHAK-----KMINTCNEYQORHGK 295  
 803 KQVERKIAQAALOGIDLDLRTVOQVQNKQKQKH 838

RESULT 39  
 AAW71295  
 ID AAW71295 standard; protein; 1312 AA.  
 XX AC AAW71295;  
 XX DT 25-NOV-1998 (first entry)  
 XX DE Human homologue of yeast RAD50.  
 XX KW Human homologue; yeast RAD50; Drosophila Septin-2; Acyl-CoA synthetase;  
 KW immunomodulatory activity; identification; activated T-cell.  
 XX OS Homo sapiens.  
 XX PN WO9838306-A1.  
 XX PD 03-SEP-1998.  
 XX PF 27-FEB-1997; 97WO-US003159.  
 XX PR 27-FEB-1997; 97WO-US003159.  
 XX PA (GENE-) GENELABS TECHNOLOGIES INC.  
 XX PI Dolganov G;  
 XX PS WPI; 1998-481207/41.  
 DR N-PSDB; AAV59979.  
 XX PT Novel human immunomodulatory poly-peptide(s) - have homology to the yeast  
 PT RAD50 or Drosophila Septin-2 proteins.  
 XX PS Disclosure; Page 136-140; 155pp; English.  
 XX CC The present sequence represents a human homologue of the yeast S.  
 CC cerevisiae gene RAD50. The present sequence has 35% overall homology to  
 CC the yeast RAD50 gene, and is expressed in activated T-cells, testis,  
 CC foetal liver and heart tissues. The specification also describes  
 CC sequences encoding human homologues of the yeast RAD50, the Drosophila  
 CC Septin-2 and Acyl-CoA synthetase. The proteins have immunomodulatory  
 CC activity. The nucleic acids and proteins can be used to identify  
 CC activated T-cells in a sample population. They can also be used to  
 CC isolate and identify sequences encoding other proteins or other compounds  
 CC having immunomodulatory activity  
 XX SQ Sequence 1312 AA;  
 Query Match 7.0%; Score 106.5; DB 2; Length 1312;  
 Best Local Similarity 18.9%; Pred. No. 15;



Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;

Qy 10 VEVVKSIAETADGALDLYNKYLQVIPWKTDFETIK-----ELSRFKQY 54  
Db 502 MEVISLQNERAD--LDRTLRKLQEMEBQLNHHHTTTRQMEMMLTKDKADKDEQIRKIKSRH 559  
Qy 55 SQEASVLVGDIKVLLMDSQDKYFEATQTVVEMCGVVTQLLSAYILLFDEYNEKKASQKD 114  
Db 560 SDELTSLLG-----YFNKKQLEDWLHSSK-----EINQTR----- 591  
Qy 115 ILIRILDGKVKLNEAQSLLTSQSFNNASGKLLALDSQLTNDPSEK-----SSYFQS 168  
Db 592 -----DRLAKLN--KELASEQNKHNINNELKREEQSS-YEDKLFVCGSQDFES 640  
Qy 169 QVDRIKE-----AYAGAAGIVAGPF----- 190  
Db 641 DLRLKEEIEKSKQRAMLAGATA--VYSQFITQLTDENQSCCPVQVQTEAELEQVI 698  
Qy 191 -----GLI-ISYSIAAGVIEGKLIPELNNRLKT 217  
Db 699 SDLSKRLAPDKLKSTESLKKKKRDEMGLVPMRQSIID--LKEKEIPELNNKLQN 756  
Qy 218 VQNFFTSLSATVQAKNDIDAAKIKLATEIAAIGETETETTT-----RFVVDY 266  
Db 757 V-----NEDIQRLKNDIEEQETLIGTIMPEESAKVCLTDVTIMERQOMEL 802  
Qy 267 DDLMLSLKGAOK-----KMINTCNEYQORHGKK 295  
Db 803 KDVERRIAQAAKLGQIDLDLORTVQVQVQKQKQK 838

RESULT 40  
ID ABB63519 standard; protein; 2346 AA.  
XX AC ABB63519;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 17349.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical.  
XX OS Drosophila melanogaster.  
XX PN WC200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US0009231.  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX PR 11-JUL-2000; 2000US-00614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX DR WPI; 2001-656860/75.  
XX DR N-PSDB; ABL07622.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX PT interactions.  
XX PS Disclosure; SEQ ID NO 17349; 2lpp + Sequence Listing; English.  
XX CC The invention relates to an isolated nucleic acid detection reagent  
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
XX CC useful in developmental biology and in elucidating cell signalling and  
XX CC cell-cell interactions in higher eukaryotes for the development of  
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at fcp.wipo.int/pub/published\_sequences  
XX Sequence 2346 AA;  
Qy 25 DLNKKYLDQVIPWKTDFETIKELS---RFKQESQEASVLVGDIKVLLMDSQDKYFEATQ 81  
Db 203 ELQNIIRREHTINTMQOSCLKTEKESLKMQEQYQAVTKTIGELTSKIEMQNDTAFKQ 262  
Qy 82 TVVEWCGVVTQLLSAYILLFDEYNEKKAS--AQKDILIRILDGKVKLNEAOK--SLLTS 137  
Db 263 ATHEYVGLKKELDKAKEKLFEPKSTESDHLIQRELLQGISIKLLLEAESQCAQLTE 322  
Qy 138 -----SQSPNNASGKLLALDSQLT--TNDF--SEKSSYFQSVQVDRIKREYAGAAAG 184  
Db 323 QMETMKQKHSAELEQNKIQAMEQELASANDLLKQARESNLESATCQLAPSA----- 375  
Qy 185 IVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQAKNDIDAAKIKLA 244  
Db 376 -----AVASRLIRSDL-----SLTELYSMYAKSSELENRNCIEIQLKLOLK 417  
Qy 245 TEIAAIGEI-----KTETETTRFVVDYDDLMLS--LKGAAKMKMINTCNE 287  
Db 418 SIATSEISAPILEKQNSDYQMKETNSELLR---EHDQLLQNKLCLERLERLALSTLNH 474  
Qy 288 YQORHGK 294  
Db 475 NQENKK 481

RESULT 41  
ID ADH62807 standard; protein; 4734 AA.  
XX AC ADH62807;  
XX DT 15-APR-2004 (first entry)  
XX DE Lactobacillus johnsonii mucin binding protein sequence SeqID25.  
XX KW mucin binding; probiotic; glycoprotein; epithelial cell;  
XX KW gastrointestinal tract; lung; uterine cervix.  
XX OS Lactobacillus johnsonii.  
XX PN EP1382970-A1.  
XX PD 21-JAN-2004.  
XX PF 15-JUL-2002; 2002EP-00015609.  
XX PR 15-JUL-2002; 2002EP-00015609.  
XX PA (NEST ) SOC PROD NESTLE SA.  
XX PI Pridmore RD;  
XX DR WPI; 2004-111519/12.  
XX DR N-PSDB; ADH62878.  
XX PT Novel mucin binding polypeptide encoded by novel gene of Lactobacillus  
XX PT johnsonii, useful for binding bacteria to mucin.  
XX PS Claim 8; SEQ ID NO 25; 225pp; English.  
XX CC This invention relates to novel mucin binding polypeptides and the DNA

CC sequences which encode them. In particular, the genes are derived from  
CC Lactobacillus johnsonii. The invention may allow a better understanding  
CC of the mechanism by which probiotics (such as L. johnsonii) can bind to  
CC mucins, large glycoproteins present on the surface of a large number of  
CC epithelial cells including those in the gastrointestinal tract, the lung  
CC or the uterine cervix. The present sequence is that of an L. johnsonii  
CC mucin binding protein of the invention.  
XX  
XX  
SQ Sequence 4734 AA;

Query Match 7.0%; Score 106.5; DB 8; Length 4734;  
Best Local Similarity 23.8%; Pred. No. 81;  
Matches 73; Conservative 38; Mismatches 109; Indels 87; Gaps 13;  
QY 2 TSIFAEQTVVVKSAIETAGALDLYN-----KYLDQV-----IPWK 38  
DB 4126 TNISSEKTDLIKATEAINAKDNINATNSVEVTAQVDGEKAIADVTGPGLSDIKKE 4185  
QY 39 TFDETIKELSRFKQYSQASVLDGDKVLLMSQDKYFEATQTVYVWCQVVTQLLSAYI 98  
DB 4186 SIDLINKALNE-KQDEINNASNLQDESTELIDQAKKI--ATEAINENNAOTN----- 4236  
QY 99 LLEFDYNEKKSAAQKDI---LIRILDGQVKKLEAKSLLTSSQ--FNNASGKLALDSQ 154  
DB 4237 ---DEAKAAQTVGVKNVSPSTEDAKKNATQIIDDALNSKQNEINNASN---LTDSE 4290  
QY 155 LTNDFSEKSYFQSQVDRIKKEAVAGAAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNR 214  
DB 4291 KTDLINQ-----ATEIANAAKDAINSATT-----NTAVEAAEYKGVADINN- 4331  
QY 215 LKTQNFSTSLSATVQKQKDI-----DAAKLKL---ATEIAAIGEI 253  
DB 4332 -----IHFTNLDDSKKAANSALIEDALTTKKDEINNASNLSDSEKAKLINQATEIANAKA 4386  
QY 254 KTEETT 260  
DB 4387 AINNAIT 4393

RESULT 42  
ABB71136  
ID ABB71136 standard; protein; 7201 AA.  
XX ABB71136;  
XX  
XX  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 40200.  
XX Drosophila melanogaster polypeptide; cell signalling; insecticide;  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX N-PSDB; ABL15239.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions.  
PT

XX Disclosure; SEQ ID NO 40200; 21pp + Sequence Listing; English.  
PS  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 7201 AA;

Query Match 7.0%; Score 106.5; DB 4; Length 7201;  
Best Local Similarity 18.9%; Pred. No. 1.4e+02;  
Matches 68; Conservative 51; Mismatches 131; Indels 109; Gaps 14;  
QY 18 ETADGA-LDLYNKYLDQVIPW-----KTDETIKEL 47  
DB 3504 QEAQGVQLSSYQDILNQTVNMLDQVEKLIHNENPASWTSQAETRSKLYKYKATNQDINSH 3563  
QY 48 SRFKQYSQASVLDGDKVLLMSQDKYFEATQTVYVWCQ-----VVTQLLSAYILL-- 100  
DB 3564 KRIVEAVNEKAAALIGSAAANADEISKAVAEVKNRYDQVQDCAKLVADLDGAFVYVQ 3623  
QY 101 PDEYNEKKSAAQKQILIRI-----LDGQVKKLEAKSLLTSSQSFNNSGKL 148  
DB 3624 FSELOKAQDYQKQLWDLTGVSYSNGKAALQARLQKINEIODAL-----PEGVAKL 3676  
QY 149 LALDSQLTNDFSEKSYFQSQVDRIKKEAVAGAAAGIVAG--PFGLLIISYSIAAGVIEGK 206  
DB 3677 KSLEHDIEQQ-----ASNIPARSKVEMADLANLHADFEKFGASLS-DVKSG-LENR 3726  
QY 207 L-----IPELNNRLKTVQ-----NFTSLSATVQKQKQIDA 238  
DB 3727 LQWNDYEINLDRLLITWLGEAENSLKNYNLKSFFEEKEQLNGFQSLAQLNRQNEADFDK 3786  
QY 239 AKLKATEIAAIGE-----IKTETETFRFYVDYDDLMLSLKGAACKMINTCNEYQQRHG 293  
DB 3787 VKDDTSELVQSGGTRIAVNVQVSSRF-----QSIQATAKEILLKCEQAVQDHG 3836  
RESULT 43  
ABB64018  
ID ABB64018 standard; protein; 685 AA.  
XX ABB64018;  
XX  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 18846.  
XX Drosophila melanogaster polypeptide; cell signalling; insecticide;  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX



PD 12-JUN-2003.  
 XX 25-NOV-2002; 2002WO-US036123.  
 PF 29-NOV-2001; 2001US-0333777P.  
 PR 18-NOV-2002; 2002US-0426742P.  
 XX (AMHP ) WYETH HOLDINGS CORP.  
 FA Fletcher LD, Memichael JC, Russell DP, Zagursky RJ;  
 PI WPI: 2003-505284/47.  
 XX N-PSDB; ADB08929.  
 DR New Alloiooccus otitidis polynucleotides and polypeptides, useful for  
 PT treating and diagnosing diseases, drug screening assays and monitoring of  
 PT effects during drug clinical trials.  
 XX Claim 33; SEQ ID NO 2870; 1019pp; English.  
 PS The present invention describes an isolated polynucleotide (I) of  
 XX Alloiooccus otitidis genomic DNA, which encodes an antigenic protein.  
 CC Alloiooccus otitidis is a Gram-positive bacterium. Also described: (1)  
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an  
 CC expression vector comprising the novel isolated polynucleotide (I), its  
 CC complement, degenerate variant or fragment; (3) a genetically engineered  
 CC host cell, transfected, transformed or infected with the vector of (2);  
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
 CC composition comprising the polypeptide, its complement, biological  
 CC equivalent or fragment, or the polynucleotide that is comprised in the  
 CC expression vector; (6) a pharmaceutical composition comprising the  
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
 CC immunising against Alloiooccus otitidis by administering to a host the  
 CC immunogenic composition; (9) detecting and/or identifying Alloiooccus  
 CC otitidis in the biological sample; (10) a kit comprising a container  
 CC containing the novel polynucleotide, its degenerate variant or fragment,  
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
 CC genetically engineered host cell under conditions suitable to produce the  
 CC polypeptide from the culture. (I) can be used in gene therapy. The  
 CC polynucleotides, polypeptides, antibodies and compositions of the present  
 CC invention can be used for treating and diagnosing diseases, drug  
 CC screening assays and monitoring of effects during drug clinical trials.  
 CC The polynucleotides are useful for expressing and detecting Alloiooccus  
 CC otitidis. The present sequence represents an Alloiooccus otitidis  
 CC antigen protein from the present invention.  
 XX Sequence 863 AA;  
 SQ  
 Query Match 7.0%; Score 106; DB 6; Length 863;  
 Best Local Similarity 20.3%; Pred. No. 9.9;  
 Matches 61; Conservative 54; Mismatches 107; Indels 78; Gaps 11;  
 QY 4 IFABQTVVVKSAIETA-----DGALDLYNKYLD-----Q 33  
 DB 460 MFSREDMEIQRLIKSSINFELFENQMVNNHNSNNEKDYLYLYKELIETSELFWD 519  
 QY 34 VIPMKTPTETIKLSR-----FKQYEQEASVLVGDIX--VLLMDSODKTF 77  
 DB 520 TITYKSNKTEIKELIKVPSITYAGKSGICLFLEAYENNLLDKLKHLMKDIYRSFI 579  
 QY 78 BATQTVYEWCVVQLSAYILLFDEYNEKASAKQDILRLDDGVKKLNEAKSKLTS 137  
 DB 580 KIVNNPYSCGIYDGTSGLFYLIY-KYQVYTRFNKNKELLISFLKIMKNVQYDKQDIMS 638  
 QY 138 SQSFNAGKLLALDSQLTDFSEKSYFOSQV-----DRIRK-----EAYAGAAG 184  
 DB 639 GQ-----AGLKLILLYLYN--GEKYDWIKEGVINQSIIDKINSNVLHKKYLYGSHG 691  
 QY 185 IVAGPFGLLIISYIAAGVIEGKLPILNLRNLTQVNTFTSLSATVQKANDIDAALKLA 244  
 DB 692 LV-----GIISVLVESTKINKIFKNIENIVENFVYILS-SVMENKGIPTIEKLS 740  
 PB  
 RESULT 46  
 AAG82283  
 ID AAG82283 standard; protein; 885 AA.  
 XX AC AAG82283;  
 XX 03-SEP-2001 (first entry)  
 DT S. epidermidis open reading frame protein sequence SEQ ID NO:1660.  
 XX S. epidermidis epidermidis SR1 strain; infection; diagnosis; vaccination;  
 DE Staphylococcus epidermidis.  
 XX Staphylococcus epidermidis.  
 OS Staphylococcus epidermidis.  
 XX WO200134809-A2.  
 PN 17-MAY-2001.  
 XX 09-NOV-2000; 2000WO-US030782.  
 XX 09-NOV-1999; 99US-0164258P.  
 PR (GLAX ) GLAXO GROUP LTD.  
 PA Kimmerly WJ;  
 XX WPI: 2001-316495/33.  
 XX N-PSDB; AAH53133.  
 DR Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 XX useful for vaccinating against infections, e.g. endocarditis.  
 PT Claim 18; Page 462; 2188pp; English.  
 PS AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II) given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
 CC and (II) can have antibacterial activity and therefore can be used in  
 CC vaccination. The nucleic acids (I) may be used to produce the S.  
 CC epidermidis polypeptides (II) via the production of vectors containing  
 CC them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH5091 to AAH5098  
 CC represent oligonucleotide sequences and primers which are used in the  
 CC exemplification of the present invention. N.B. The present invention  
 CC specifically claims all the polynucleotide sequences given in the  
 CC sequence listing of the present specification, however the sequence  
 CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
 CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
 CC for SEQ ID NO:4455 to 4464  
 XX Sequence 885 AA;  
 SQ  
 Query Match 7.0%; Score 106; DB 4; Length 885;  
 Best Local Similarity 18.5%; Pred. No. 10;  
 Matches 54; Conservative 51; Mismatches 111; Indels 76; Gaps 8;  
 QY 42 ETIKELSFQKQYEQEASVLVGDILKVLMDSDQKYFEATQTVYEMC----- 87  
 DB 2 RPLKEEAIAIKYKQLSKEMEQSDVIVTVSDIDHYEDNQRDLRLNHLKSSQAEKGGQ 61  
 QY 88 GVVTTOLLSAYILLFDEYNEKASAKQDILRLDDGVKKLNEAKSKLTSQSFNNAAGK 147  
 DB 62 AQINOLLQKY-----KGKQQN-----DYDIEKLN---YELVKATENYEQLSGK 102  
 QY 148 LLALDSQTLND-----FSEKSSYFOSQVDRIRKEAYAGAAGVAGPFGLLIISYIAAG 201  
 PB



CC The specification describes a method for directing integration of a  
CC nucleic acid of interest to a pre-determined site, where the nucleic acid  
CC has homology at or around the pre-determined site, in a eukaryote with a  
CC preference for non-homologous recombination. The method comprises  
CC steering an integration pathway towards homologous recombination. The  
CC method is useful for directing integration of a nucleic acid of interest  
CC to a subtelomeric and/or telomeric region in an eukaryote with a  
CC preference for non-homologous recombination. The nucleic acid of interest  
CC comprises an inactive gene to replace an active gene, or vice versa, is a  
CC portion of a gene delivery vehicle, confers a desired property to the  
CC eukaryotic cell, or encodes a therapeutic proteinaceous substance or a  
CC substance conferring resistance for an antibiotic substance to a cell.  
CC The method is useful for improving gene targeting efficiency. The method  
CC is useful in the replacement of an active gene by an inactive gene, for  
CC e.g. for the inactivation of genes controlling undesired side branches of  
CC metabolic pathways, to increase the quality of bulk products such as  
CC starch, or to increase the production of specific secondary metabolites  
CC or to inhibit formation of unwanted metabolites, and also to inactivate  
CC genes controlling senescence in fruits and flowers or that determine  
CC flower pigments. The method is also useful for replacing an inactive gene  
CC by an active gene. For e.g. the replacement of a defective p53 by an  
CC intact p53. Many tumours acquire a mutation in p53 during their  
CC development which renders it inactive and often correlates with a poor  
CC response to cancer therapy. By replacing the defect p53 by an intact p53,  
CC e.g. through gene therapy, conventional anti cancer therapy have better  
CC changes of succeeding. The method is also useful for therapeutic  
CC proteinaceous substance integration. A tumoricidal gene can be delivered  
CC to a pre-determined site present only in e.g. proliferating cells, or  
CC present only in tumour cells, e.g. to the site where a tumour antigen is  
CC expressed form. ABB77984-86 represent RAD50 homologues. RAD50 is involved  
CC in non-homologous recombination

XX Sequence 1318 AA;

Query Match 7.0%; Score 105.5; DB 5; Length 1318;  
Best Local Similarity 18.4%; Pred. No. 19;  
Matches 72; Conservative 53; Mismatches 106; Indels 161; Gaps 14;

QY 10 VEVVKSATETADGALDLYNKYLDQVTPWKTFDETIK-----ELSRFKQY 54  
DB 508 MEVISLQNEKAD--LDRTLRLKLDQEMEQNLNHHHTTTRTQEMELTKADKADQEKIRKSRH 565  
QY 55 SQEASVILVGDIKVLLMDSQDKYFATQVYEWCGVVTQLLSAYILLPFDYNEKKASAKQD 114  
DB 566 SDELTSLLG-----YFPNKKQLEDWLHLSKSK-----EINQTR----- 597  
QY 115 ILIRILDGCVKKLNEAKSKLTSQSFNNASGKLALDLSQLTNDFSEK-----SSYFQS 168  
DB 598 -----DRLAKLN---KELASSEQNKHNINNELERKEQQLSS--YEDKLFVDCGSDPFES 646  
QY 169 QVDRIRKE-----AVAGAAA-----GIVAGPFLGIISYSIAAGV 202  
DB 647 DLDRLKEEIEKSKQRAMLAGATVYSQITQLTNDENQCCPQCVQRVQTEAELQEAISD 706  
QY 203 IEKGL-----IPELNNRLTKVQNF 221  
DB 707 LQSKLRLAPDLKSTSELKXKXKRDMLGLAPMRQSIIDLKEKEIPELNRKLNQV--- 763  
QY 222 FTSLSATVQANKDIAAKKLKATEIAAIGETIKTETTT-----RFYVDYDDILM 270  
DB 764 -----NRDIOQLRNDIEEQETLLGTTMPDEESAKVCLTDTVTIMERFQEMELKQVE 812  
QY 271 LSLKGAAK-----KMINTCNEYQQRHGKK 295  
DB 813 RKTAQAQAKLQIGDLDRLTRVQVNOEKQEKQHK 844

RESULT 49  
ADJ68860  
ID ADJ68860 standard; protein; 1318 AA.

XX  
AC ADJ68860;

06-MAY-2004 (first entry)  
XX Human heat mitochondrial protein as a therapeutic target SeqID666.  
DE  
XX  
XX Mitochondrial; human; screening assay; diabetes mellitus;  
KW Huntington's disease; osteoarthritis; LHON;  
KW Leber's hereditary optic neuropathy; LHON;  
KW Mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
KW neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;  
KW osteopathic; ophthalmological; cytostatic.

OS Homo sapiens.

XX WO2003087768-A2.

XX 23-OCT-2003.

XX 04-APR-2003; 2003WO-US010870.

XX 12-APR-2002; 2002US-0372843P.

XX 17-JUN-2002; 2002US-0389987P.

XX 20-SEP-2002; 2002US-0412418P.

XX (MITO-) MITOKOR.

XX (BUCK-) BUCK INST AGE RES.

XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;  
PI Warnock DE;

XX WPI; 2003-845369/78.

XX Identifying a mitochondrial target for drug screening assays and for  
PT treating diseases associated with altered mitochondrial function,  
PT comprises detecting a modified polypeptide in a sample and correlating  
PT with the disease.

XX Claim 1; SEQ ID NO 666; 180pp; English.

XX This invention relates to novel mitochondrial targets that can be used  
CC for therapeutic intervention in treating a disease associated with  
CC altered mitochondrial function. Specifically, it refers to a method for  
CC identifying proteins of the human heart mitochondrial proteome that are  
CC useful for drug screening assays, as well as therapeutic targets. The  
CC present invention describes a method for identifying such proteins that  
CC can be used in the treatment of various diseases associated with altered  
CC mitochondrial function including diabetes mellitus, Huntington's disease,  
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
CC compositions have neuroprotective, nontropic, antidiabetic,  
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
CC cytostatic activities. This polypeptide sequence is a human heart  
CC mitochondrial protein of the invention.

XX Sequence 1318 AA;

Query Match 7.0%; Score 105.5; DB 7; Length 1318;  
Best Local Similarity 18.4%; Pred. No. 19;  
Matches 72; Conservative 53; Mismatches 106; Indels 161; Gaps 14;

QY 10 VEVVKSATETADGALDLYNKYLDQVTPWKTFDETIK-----ELSRFKQY 54  
DB 508 MEVISLQNEKAD--LDRTLRLKLDQEMEQNLNHHHTTTRTQEMELTKADKADQEKIRKSRH 565  
QY 55 SQEASVILVGDIKVLLMDSQDKYFATQVYEWCGVVTQLLSAYILLPFDYNEKKASAKQD 114  
DB 566 SDELTSLLG-----YFPNKKQLEDWLHLSKSK-----EINQTR----- 597  
QY 115 ILIRILDGCVKKLNEAKSKLTSQSFNNASGKLALDLSQLTNDFSEK-----SSYFQS 168  
DB 598 -----DRLAKLN---KELASSEQNKHNINNELERKEQQLSS--YEDKLFVDCGSDPFES 646

Qy 169 QVDRIRKE-----AYAGARA-----GIVAGPFGLLISYSIAAGV 202  
Db 647 DLRLKEIEKSKQRAMLAGATAVYSQFTQLTDENQSCCPVCQVQFQTEAELOEALSD 706  
Qy 203 IEGKL-----IPELNNRLKTVQNF 221  
Db 707 LQSKRLAPDKLKSTESSELKKKRRDEMGLAPMQSIIDLKEKEIPELRNKLQNV--- 763  
Qy 222 FTLSATVQKANKDIDAAKLKLATEIAAIGETETETTT-----RFVYDDLM 270  
Db 764 -----NRDIQRLKNDIEQETLLGTIMPEESAKVCLTDVTIMERFQMEKLDVE 812  
Qy 271 LSLKGAAG-----KMINTCNEYQORHGKK 295  
Db 813 RKTAQQAQKLGIDLDRTVQVQVQKQKQHK 844

RESULT 50

ADJ66509  
ID ADJ66509 standard; protein; 1318 AA.  
XX AC ADJ66509;  
XX DT 06-MAY-2004 (first entry)  
XX DE RAD50 homolog HSRAD50 for anti-cancer protein complex.  
XX KW neuroprotective; cytostatic; gene therapy; protein complex;  
XX OS cellular network; cancer; neurodegenerative disease; drug target.  
XX OS Homo sapiens.  
XX PN WO2004009622-A2.  
XX PD 29-JAN-2004.  
XX PF 18-JUL-2003; 2003WO-EP007835.  
XX PR 19-JUL-2002; 2002EP-00016109.  
XX PR 19-JUL-2002; 2002EP-00016111.  
XX PR 19-JUL-2002; 2002EP-00016123.  
XX PR 19-JUL-2002; 2002EP-00016128.  
XX PR 22-JUL-2002; 2002EP-00016427.  
XX PA (CELL-) CELLZONE AG.

XX Merino A, Bouwmeester T, Bauer A, Drewes G, Marzloch M, Kruse U;  
XX Superti-Furga G, Eberhard D, Ruffner H, Hobson S, Helftenbein G;  
XX Cruciat C;  
XX WPI; 2004-123372/12.

XX New protein complexes of cellular networks underlying the development of  
XX cancer and other diseases, useful for diagnosing and/or treating  
XX neurodegenerative diseases or cancer, and in drug screening.

XX Disclosure; SEQ ID NO 39; 809pp; English.

XX The invention relates to a protein complex of cellular networks  
XX underlying the development of cancer and other diseases. The complex (I)  
XX comprises at least one first and second proteins selected from any of the  
XX proteins listed in the specification, or their functionally active  
XX derivatives, fragments, homologues or variants, the variants being  
XX encoded by a nucleic acid that hybridizes to the nucleic acid encoding  
XX the protein under low stringency conditions. A complex (II) comprises at  
XX least two of the second proteins, where the low stringency conditions  
XX comprise hybridization in a buffer comprising 35% formamide, 5 x SSC, 50  
XX mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% BSA, 100 microg/ml  
XX denatured salmon sperm DNA, and 10% (wt/vol) dextran sulfate for 18-20  
XX hours at 40 deg C, washing in a buffer consisting of 2 x SSC, 25 mM Tris-  
XX HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS for 1.5 hours at 55 deg C, and  
XX washing in a buffer consisting of 2 x SSC, 25 mM Tris-HCl (pH 7.4), 5 mM  
XX EDTA, and 0.1% SDS for 1.5 hours at 60 deg C. The composition and methods

CC are useful in diagnosing or treating diseases and disorders, preferably  
CC neurodegenerative diseases. These may also be used as a drug target or in  
CC manufacturing a medicament for the treatment or prevention of the above-  
CC mentioned diseases or disorders. The composition may also be used for  
CC treating cancer. This sequence represents one of the proteins of the  
CC complex of the invention.

SQ Sequence 1318 AA;

Query Match 7.0%; Score 105.5; DB 8; Length 1318;  
Best Local Similarity 18.4%; Pred. No. 19;  
Matches 72; Conservative 53; Mismatches 106; Indels 161; Gaps 14;

Qy 10 VEYVKSATETADGALDLYNKYLDVIVPKWTFDTIK-----ELSRFQOEY 54  
Db 508 MEVISLQNEKAD--LDRTLKLDQEMEQMLHHTTTTQMEMLTKDKADKDEQIKIKSRH 565  
Qy 55 SQEASVLVGBDIKVLMLDSQDKYFEATQTVYEWGCVTQLLSAVILLPDEYNEKKASAKQD 114  
Db 566 SDELTSLG-----YFPNKKQLEDWLHLSKSK-----EINQTR----- 597  
Qy 115 ILIRIILDDGVKLNKAEAKSLTSSQSPNNSAGKLLALDSQLTNDPSEK-----SSYFQS 168  
Db 598 -----DRLAKLN--KELASSEQNKHNINNELERKEEQLSS-YEDKLPDVCQSQDFES 646  
Qy 169 QVDRIRKE-----AYAGAAA-----GIVAGPFGLLISYSIAAGV 202  
Db 647 DLRLKEIEKSKQRAMLAGATAVYSQFTQLTDENQSCCPVCQVQFQTEAELOEALSD 706  
Qy 203 IEGKL-----IPELNNRLKTVQNF 221  
Db 707 LQSKRLAPDKLKSTESSELKKKRRDEMGLAPMQSIIDLKEKEIPELRNKLQNV--- 763  
Qy 222 FTLSATVQKANKDIDAAKLKLATEIAAIGETETETTT-----RFVYDDLM 270  
Db 764 -----NRDIQRLKNDIEQETLLGTIMPEESAKVCLTDVTIMERFQMEKLDVE 812  
Qy 271 LSLKGAAG-----KMINTCNEYQORHGKK 295  
Db 813 RKTAQQAQKLGIDLDRTVQVQVQKQKQHK 844

Search completed: January 5, 2005, 10:56:28

Job time : 70.7089 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:51:49 ; Search time 16.0526 Seconds  
(without alignments)  
1260.042 Million cell updates/sec

Title: US-09-993-292B-2  
Perfect score: 1515  
Sequence: 1 MTSIPAEQTVVVKSALETA.....NEYQQRHGKTLFEPDVAS 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS-COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 1396  | 92.1        | 309    | 1     | Sequence 3, Appli |
| 2          | 1396  | 92.1        | 309    | 5     | Sequence 3, Appli |
| 3          | 121.5 | 8.0         | 1036   | 4     | Sequence 7736, Ap |
| 4          | 114   | 7.5         | 515    | 4     | Sequence 5317, Ap |
| 5          | 112.5 | 7.4         | 722    | 4     | Sequence 20613, A |
| 6          | 111   | 7.3         | 1231   | 4     | Sequence 5150, Ap |
| 7          | 108.5 | 7.2         | 1066   | 3     | Sequence 8, Appli |
| 8          | 108.5 | 7.2         | 1066   | 4     | Sequence 8, Appli |
| 9          | 108.5 | 7.2         | 1066   | 4     | Sequence 8, Appli |
| 10         | 108.5 | 7.2         | 3696   | 3     | Sequence 5080, Ap |
| 11         | 107   | 7.1         | 1211   | 3     | Sequence 4820, Ap |
| 12         | 106.5 | 7.0         | 1312   | 2     | Sequence 148, App |
| 13         | 106.5 | 7.0         | 1312   | 2     | Sequence 51, Appl |
| 14         | 106.5 | 7.0         | 1312   | 4     | Sequence 148, App |
| 15         | 106   | 7.0         | 815    | 4     | Sequence 4284, Ap |
| 16         | 106   | 7.0         | 885    | 4     | Sequence 1660, Ap |
| 17         | 105.5 | 7.0         | 924    | 4     | Sequence 18798, A |
| 18         | 104.5 | 6.9         | 566    | 4     | Sequence 807, App |
| 19         | 104.5 | 6.9         | 1196   | 4     | Sequence 3944, Ap |
| 20         | 103.5 | 6.8         | 718    | 4     | Sequence 2753, Ap |
| 21         | 103   | 6.8         | 606    | 4     | Sequence 2, Appli |
| 22         | 103   | 6.8         | 631    | 4     | Sequence 11, Appl |
| 23         | 102   | 6.7         | 1976   | 4     | Sequence 1078, Ap |
| 24         | 101   | 6.7         | 477    | 1     | Sequence 3, Appli |
| 25         | 101   | 6.7         | 477    | 1     | Sequence 3, Appli |
| 26         | 101   | 6.7         | 477    | 3     | Sequence 3, Appli |
| 27         | 101   | 6.7         | 1939   | 3     | Sequence 1, Appli |

Query Match 92.1%; Score 1396; DB 1; Length 309;

#### ALIGNMENTS

##### RESULT 1

US-08-557-115-3  
; Sequence 3, Application US/08557115  
; Patent No. 5731151  
; GENERAL INFORMATION:  
; APPLICANT: King, Harold C.  
; APPLICANT: Sathish, Mundayoor  
; APPLICANT: Shinnick, Thomas M.  
; TITLE OF INVENTION: Regulator of Contact-Mediated Hemolysis  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jones & Askew  
; STREET: 191 Peachtree Street, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.A.  
; ZIP: 30303

##### COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 26-JUN-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Greene, Jamie L.  
REGISTRATION NUMBER: 32,467  
REFERENCE/DOCKET NUMBER: 03063-0171US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 818-3700  
TELEFAX: (404) 818-3799  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 309 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-557-115-3

28 101 6.7 1939 4 US-09-538-092-917 Sequence 917, App  
29 100.5 6.6 643 4 US-09-538-092-844 Sequence 844, App  
30 100.5 6.6 1454 4 US-09-538-352-5793 Sequence 5793, Ap  
31 100 6.6 1960 4 US-09-538-092-1077 Sequence 1077, Ap  
32 100 6.6 2871 4 US-09-538-092-936 Sequence 936, App  
33 100 6.6 10182 3 US-09-134-001C-3159 Sequence 3159, Ap  
34 98.5 6.5 1010 3 US-09-134-001C-5178 Sequence 5178, Ap  
35 98 6.5 284 4 US-09-134-000C-3594 Sequence 3594, Ap  
36 98 6.5 733 4 US-09-328-352-5599 Sequence 5599, Ap  
37 98 6.5 1886 3 US-08-938-105-3 Sequence 3, Appli  
38 97.5 6.4 669 4 US-09-252-991A-31794 Sequence 170, App  
39 97.5 6.4 759 3 US-09-199-637A-170 Sequence 170, App  
40 97.5 6.4 862 4 US-09-248-796A-16013 Sequence 16013, A  
41 97.5 6.4 1857 4 US-09-917-254-91 Sequence 91, Appl  
42 97.5 6.4 1972 4 US-09-538-092-1084 Sequence 1084, Ap  
43 97.5 6.4 2285 3 US-09-308-375-2 Sequence 2, Appli  
44 97.5 6.4 2482 1 US-08-328-254-6 Sequence 6, Appli  
45 97.5 6.4 3210 4 US-09-538-092-1154 Sequence 1154, Ap  
46 97.5 6.4 3248 1 US-08-353-700-1 Sequence 1, Appli  
47 97.5 6.4 3248 5 PCT-US95-16216-1 Sequence 1, Appli  
48 96.5 6.4 961 4 US-09-914-259-66 Sequence 66, Appl  
49 96.5 6.4 350 4 US-09-489-039A-8352 Sequence 8352, Ap  
50 96.5 6.4 457 4 US-08-924-629C-6 Sequence 6, Appli

Best Local Similarity 91.4%; Pred. No. 3e-130; Matches 275; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MTSIFABQTVVVKSALETADGALDLYNKYLDQVDPWKTFTDETIKELSRFKQBYSEASV 60  
Db 1 MTEIVADKTVEVKNALLETADGALDLYNKYLDQVDPWKTFTDETIKELSRFKQBYSEASV 60

Qy 61 LVGDIKVLMDSDQKYPEATQTVVWCGVATQLLAYILLFDEYNEKKSAAQKDILIRIL 120  
Db 61 LVGDIKVLMDSDQKYPEATQTVVWCGVATQLLAYILLFDEYNEKKSAAQKDILIRIL 120

Qy 121 DDGKKVLEAKSLTSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSOVDRIKREAYAG 180  
Db 121 DDGTTKLEAKSLVSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSOVDRIKREAYAG 180

Qy 181 AAAGVAGPFGGLIISYSIAAGVIEGKLIPELNRLKTVQNFFTLSATVKQANKDIDAAK 240  
Db 181 AAAGVAGPFGGLIISYSIAAGVIEGKLIPELNRLKTVQNFFTLSATVKQANKDIDAAK 240

Qy 241 LKLAETIAAIGETETETTFYVYDDMLSLKGAAKWINTCNEYQORHGKKTLEFV 300  
Db 241 LKLAETIAAIGETETETTFYVYDDMLSLKGAAKWINTCNEYQORHGKKTLEFV 300

Qy 301 P 301  
Db 301 P 301

RESULT 2  
PCT-US94-05869-3  
; Sequence 3, Application PC/TUS9405869  
; GENERAL INFORMATION:  
; APPLICANT: King, C. H.  
; APPLICANT: Sathish, Mundayoor  
; APPLICANT: Shinnick, Thomas M.  
; TITLE OF INVENTION: REGULATOR OF CONTACT-MEDIATED HEMOLYSIS  
; TITLE OF INVENTION: OF MYCOBACTERIUM TUBERCULOSIS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Suite 1200, The Candler Building  
; STREET: 127 Peachtree Street, N.E.  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303-1811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/05869  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perryman, David G.  
; REGISTRATION NUMBER: 33,438  
; REFERENCE/DOCKET NUMBER: 1414.611  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 688-0770  
; TELEFAX: (404) 688-9880  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 309 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US94-05869-3

Query Match 92.1%; Score 1396; DB 5; Length 309;  
Best Local Similarity 91.4%; Pred. No. 3e-130;

Matches 275; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MTSIFABQTVVVKSALETADGALDLYNKYLDQVDPWKTFTDETIKELSRFKQBYSEASV 60  
Db 1 MTEIVADKTVEVKNALLETADGALDLYNKYLDQVDPWKTFTDETIKELSRFKQBYSEASV 60

Qy 61 LVGDIKVLMDSDQKYPEATQTVVWCGVATQLLAYILLFDEYNEKKSAAQKDILIRIL 120  
Db 61 LVGDIKVLMDSDQKYPEATQTVVWCGVATQLLAYILLFDEYNEKKSAAQKDILIRIL 120

Qy 121 DDGKKVLEAKSLTSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSOVDRIKREAYAG 180  
Db 121 DDGTTKLEAKSLVSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSOVDRIKREAYAG 180

Qy 181 AAAGVAGPFGGLIISYSIAAGVIEGKLIPELNRLKTVQNFFTLSATVKQANKDIDAAK 240  
Db 181 AAAGVAGPFGGLIISYSIAAGVIEGKLIPELNRLKTVQNFFTLSATVKQANKDIDAAK 240

Qy 241 LKLAETIAAIGETETETTFYVYDDMLSLKGAAKWINTCNEYQORHGKKTLEFV 300  
Db 241 LKLAETIAAIGETETETTFYVYDDMLSLKGAAKWINTCNEYQORHGKKTLEFV 300

Qy 301 P 301  
Db 301 P 301

RESULT 3  
US-09-543-681A-7736  
; Sequence 7736, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 7736  
; LENGTH: 1036  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-7736

Query Match 8.0%; Score 121.5; DB 4; Length 1036;  
Best Local Similarity 20.7%; Pred. No. 0.0068;  
Matches 80; Conservative 66; Mismatches 126; Indels 115; Gaps 17;

Qy 2 TSIFAEQTVBWK-----SAIETADGALDLYNKYLDQVDPWKTFTDETIKELSRFKQ 53  
Db 167 TALSSMEINELIKKQNGENVSSSELAKASIDILINQLVD-----TASSLNNNISAFSQ 220

Qy 54 YSQEASV-----VGDIKVLLMSQDKYFEATQTVVWCGVATQLLAYILLFDEY 104  
Db 221 LNKLSVLSNTKHLNGVGN-KLQNLPLNDKLTGLDVT-----SGLSAISASFILSNADAD 276

Qy 105 NEKASAAQKOLIRILDGVKKLNE-----AOKSLTSSQSFNNASGKLL 149  
Db 277 TGTAAAGVELTITKLVGNVGVKASQYILAQORVAGLSTSAASAGLIASAVTLAISPLSFL 336

Qy 150 ALDSQ-----TNDPSEKSSYFQ-----SQVDRIKRE-----YAGAAA 183  
Db 337 AIADQFKANKIEEVSQRFKKFGYEGDLSLAAPFKETGAIDATTTINTALGTISAGISA 396

Qy 184 GIVAGPFGGLIISYSIAA--GVIEG-----KLPELNRLKTV-----QNPTS- 224  
Db 397 ASTASLIGAPISALVGAITGIISGILKASQSMFHVANRWANTIAEWKTHGNFPENG 456

Qy 225 -----LSATVK---QANKDIDAAKIKLAT-----ETAAIGETETETTFYV 264

[illegible]

```

176 176 ---BAYAGAAGIVAGPGLIIISYIAAGVIEGKLIPELNRLKTVQNFPTSLSATVQQA 232
230 230 RGYEYIRSQGEVRDG-----VTTVNAAAVPEEVIGSEVFDLKRS--NYNLAQYATVKTV 282
233 233 NKDIDAAKLIKLAT-----EIAAIGEIKTETETTRFVVDY----- 266
283 283 SN--GQKYPVATVQQAVLATKABLAELGDDIAEMFTS--VDYKVEVTRAGKIALSNEVV 337
267 267 DDLMLSLKGA-----AKQWINTCNEY-----QQRHGKTKTLFEVDPDV 303
338 338 EDSAVNIVQEVKQOLAKVLVENTDNKHIMDLLKTFKTKTAATLDDL 382

RESULT 5
US-09-248-796A-20613
; Sequence 20613, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDID
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20613
; LENGTH: 722
; TYPE: PRNT
; ORGANISM: Candida albicans
US-09-248-796A-20613

```

|                       |                  |  |            |             |
|-----------------------|------------------|--|------------|-------------|
| Query Match           | 7.4%             | Score 112.5;   | DB 4;      | Length 722; |
| Best Local Similarity | 22.2%;           | Pred. No. 0.031;   |            |             |
| Matches 58;           | Conservative 40; | Mismatches 90;   | Indels 73; | Gaps 10;    |
| Qy                    | 38               | KTPEDETIKLSRPFKQEQSASVLVGDIKVLLMDSQDKYFEATQ-----TVYEWCGVVTQ  | 92         |             |
| Db                    | 218              | KTLKSTREELNGSKTEILR-----LKALLRESEDELYQVKQENYKTSVHDYEQDLAQ    | 269        |             |
| Qy                    | 93               | L-----LSAYILLDFEYNEKKAQAQKDILI--RILDDGVKKLNEAQS              | 133        |             |
| Db                    | 270              | LKVKHETLLSRNKDINESLEIYKKRSDEYKYKUELAESAIAISKRHEEQATKEMKESRSQ | 329        |             |
| Qy                    | 134              | LLTSQSQFNNAAGKLLALDSQLDNDFSEKSYFQSQVDRIKRAYAGAAAGIVAGPFLGI   | 193        |             |
| Db                    | 330              | LLLVRBE-----LRTTQILIKDFRIKVENLEATIE-----                     | 360        |             |
| Qy                    | 194              | ISVSIAGVIEGKLIPE-LNNRKTQVN--FFTSLSATVKQAOKDIDAAKKLATEIAAI    | 250        |             |
| Db                    | 361              | KNQLDANKKEIKQIQDKLNLKFNFKNELNEKLKEEIKNLNRDLQ-----FKTDI----   | 412        |             |
| Qy                    | 251              | GEIKTETETTRFYVDYDDLML  | 271        |             |
| Db                    | 413              | -ETKLIKENKKLOLDYEDVILL                                       | 432        |             |

RESULT 6  
US-09-107-532A-5150  
; Sequence 5150, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street

|                       |        |                  |                 |                     |
|-----------------------|--------|------------------|-----------------|---------------------|
| Query Match           | 7.5%;  | Score 114;       | DB 4;           | Length 515;         |
| Best Local Similarity | 23.5%; | Pred. No. 0.013; |                 |                     |
| Matches               | 81;    | Conservative 61; | Mismatches 107; | Indels 96; Gaps 19; |

  

|    |     |  |     |
|----|-----|--|-----|
| Qy | 30  | YLQVIVPWKTFDETIKELS-----RPKQESQSE-ASVLGCDIRKVLWMSQDKYFEATQT  | 82  |
| Db | 63  | FVEVITP-----EALKEVDLSNVELLYGHDSYKPLASVAGTGLK-LNVDDTGLHFEAELT | 116 |
| Qy | 83  | -----VYEWG--GVVTQLLSAYILLFDEYNKKASQAQDILRIILDDGVGKKLNEAQS    | 133 |
| Db | 117 | DTTYSNDVYENISKGVDVDSMSFGVGLGDSFDKK-----EDGTTIERSIDIKALNEI--S | 169 |
| Qy | 134 | LLT-----SSQSFNNASGKLLALDSQI--TNDPSEKSSYPQSQVDRIRK-           | 175 |
| Db | 170 | VTTVPAYDSNNVQNKESYSFMSFNNOAKOTNNLESTSKAQESNNNEKTLIDNEKTEL    | 229 |

CITY: Walham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariandelio, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 5150:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1231 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...1231  
SEQUENCE DESCRIPTION: SEQ ID NO: 5150:  
US-09-107-532A-5150

Query Match 7.3%; Score 111; DB 4; Length 1231;  
Best Local Similarity 23.3%; Pred. No. 0.098;  
Matches 78; Conservative 58; Mismatches 125; Indels 74; Gaps 17;  
QY 6 ABOQVVKSAJETADGALDLYNK-----YLDQVVPKMTFDE-----TIKELS 48  
DB 269 ABEIVSMGSAQTIETQIDYLNANGRTGFLNHLXRPFPVETVQKLPHTVKTIAVLD 328  
QY 49 RFQKQVQSASVLVGDIVKLLMDSQ-----DKYFEATQVYEWCGVVTQLLSAYILLF 101  
DB 329 RSKEPGAGGEPILL-DVQSAIYDSELRPAVIGRYGLGSK-----DVTPDOISA---VF 378  
QY 102 DEYNEKASAKOILIRILDGKVLKNEAKQSL-LTSSQSFN-----NASGKL 148  
DB 379 DELKQPSIRKRTGFTIGIVDDVYQSLPEKESLDLTPQTFOAKFWGFGSDGTVCANKSA 438  
QY 149 LALDSQLTNDFSKSSYFQSQ-----VDRIR-----KEAVAGAAAGIVA---GPFGLII 194  
DB 439 IKIIGDHTDKYAGQYFYDYSKSGGLTVSHLRFGDTPIRSAYLVEHADIVACHTP-AYLH 497  
QY 195 SYSIAGVIEGKLIPELNNELKTQVFFTSLSATVKQ--ANKDI---DAAKLKLATEIAA 249  
DB 498 SYDLVRLKPGGGTF--LLNTLWSEDEQIETHPLPKLARYLAENNRIFYTINAMRLAQEVGL 555  
QY 250 IGIKETEETRF-----YVDYDDLMLSLKGAAKK 280  
DB 556 GRRINTAMETAFFKLADIIPFDE-VLPLLKEALK 589

RESULT 7  
US-09-541-782-8  
; Sequence 8, Application US/09541782  
; Patent No. 6284480  
; GENERAL INFORMATION:

; APPLICANT: Nislow, Corey  
; APPLICANT: Sakowicz, Roman  
; APPLICANT: Beraud, Christophe  
; TITLE OF INVENTION: Antifungal Assay  
; FILE REFERENCE: 1015  
; CURRENT APPLICATION NUMBER: US/09/541,782  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 1066  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-541-782-8  
Query Match 7.2%; Score 108.5; DB 3; Length 1066;  
Best Local Similarity 21.6%; Pred. No. 0.14;  
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;  
QY 24 LDLYNKYLDQVVPKMTFDETIKELSRFQKQVQSASVLVGDIVKLLMDSQDKYPEATQTV 83  
DB 159 LELYNEELCDLL---STDDTTK--IRIFDDSTKGSVLIQGLEIPIVHSKDDVYKLEKG 213  
QY 84 YEWCGVVTQLLSAYILLFDEYNEKASAKOILIRILDGKVK-----KLNBAQKSLTSS 138  
DB 214 KERRKTATILMNA-----QSSRSHTVFSIVVHIRENGIEGEDMLKIGKLNVLVDLAGS 265  
QY 139 QSFNNASGK-----LLALDSQLTN--DFSEKSSYFQSDVDRIKKEAYAGAA 182  
DB 266 ENVKAGNEKIRVRETVMNQSLTLGRVITALVDRAHPVYRESKLTLLQESLGRT 325  
QY 183 -----AGIVAGPFL-----IISYSIAAGVIEGKLIPELNNEL--KTQVNFFTLSATVK 230  
DB 326 KTSIIATISPGHKDIEETLSTLEYAHRKXNQK--PEVNOQLTKTVLKEYTE---EID 380  
QY 231 QANKDIDAAKLKLATEIA--AIGEIKTETETTFYVVDYDDLMLSLK 275  
DB 381 KLKRDLMWAARDKNGIYLAETVGEITLKLESONREINERKMLLKALK 427

RESULT 8  
US-09-723-820-8  
; Sequence 8, Application US/09723820  
; Patent No. 6468760  
; GENERAL INFORMATION:  
; APPLICANT: Nislow, Corey  
; APPLICANT: Sakowicz, Roman  
; APPLICANT: Beraud, Christophe  
; TITLE OF INVENTION: Antifungal Assay  
; FILE REFERENCE: 1015  
; CURRENT APPLICATION NUMBER: US/09/723,820  
; CURRENT FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 09/541,782  
; PRIOR FILING DATE: 2000-04-03  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 1066  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-723-820-8  
Query Match 7.2%; Score 108.5; DB 4; Length 1066;  
Best Local Similarity 21.6%; Pred. No. 0.14;  
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;

QY 24 LDLYNKYLDQVVPKMTFDETIKELSRFQKQVQSASVLVGDIVKLLMDSQDKYPEATQTV 83  
DB 159 LELYNEELCDLL---STDDTTK--IRIFDDSTKGSVLIQGLEIPIVHSKDDVYKLEKG 213  
QY 84 YEWCGVVTQLLSAYILLFDEYNEKASAKOILIRILDGKVK-----KLNBAQKSLTSS 138  
DB 214 KERRKTATILMNA-----QSSRSHTVFSIVVHIRENGIEGEDMLKIGKLNVLVDLAGS 265

|    |     |   |     |
|----|-----|---|-----|
| Qy | 139 | QSPFNASGK-----LIALDSQLTN--DFSEKSSYFQSOVDRIKEAVAGAA          | 182 |
| Db | 266 | ENVSKAGNEKGRVRETNIQSLTTLGERTVITVLDRAHPVPTRESKTLRLQESLGGRT   | 325 |
| Qy | 183 | -----AGIVAGPFL-----IISYSIAAGVTGKGLIPELANRL--KTVQNFFTSLSATVK | 230 |
| Db | 326 | KTSIIATSPGHKEETLTSLTVAHRAKNIQNK--PEVNQKTKTKTLVKEYTE---EID   | 380 |
| Qy | 231 | QANKOIDAAKLATEIA--AIGSIKTTETTRFYVDYDILMLSLK                 | 275 |
| Db | 381 | KLKRDLMARDKNGYGLAEETVGYHTLKLESONRELKMLLKALK                 | 427 |

```

RESULT 9
US-10-270-085-8
; Sequence 8, Application US/10270085
; Patent No. 6627408
; GENERAL INFORMATION:
;
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/10/270,085
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/723,820
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-270-085-8

```

| Query Match           | 7.2%;            | Score 108.5;  | DB 4;      | Length 1066; |
|-----------------------|------------------|---|------------|--------------|
| Best Local Similarity | 21.6%;           | Pred. No. 0.14;   |            |              |
| Matches 63;           | Conservative 62; | Mismatches 110;   | Indels 53; | Gaps 12;     |
| Qy                    | 24               | LDLYNKYLDQVWPWKTDETIKELSRFQKSYQSEASVLVDGIKVLMLDSODKYFEATQTV | 83         |              |
| Db                    | 159              | LEYANEELCDLL---STDDTTK--IRIFDSDTKSGVIOGLEBIPVHSXDDVYKLEKG   | 213        |              |
| Qy                    | 84               | YEWGVVVTQLLSAYILLFDEYNEKKAQAQDILIRILDDGVK-----KLENAQKSLTSS  | 138        |              |
| Db                    | 214              | KERKTATTLLMNA-----QSRSHTVFSVVHIRENGIEGEDMLUKIKLNILNVLDA     | 265        |              |
| Qy                    | 139              | QSFNNASGK-----LLALDLSOLTN--DFSEKSSYFSQSDVRIRKEAYAGAA        | 182        |              |
| Db                    | 266              | ENVSKAGNEKGIURVETWVNIQSLLTLGRVITALVDRAHPVPYRESKLTLLQSSIGGRT | 325        |              |
| Qy                    | 183              | -----AGIVAGPGL-----IISYSAAGVIEGKLIPELNNRL--KTQVHFPSLSATVK   | 230        |              |
| Db                    | 326              | KTSLIATISPGHKDIBETLSLEYAHRAKQIQNK--PEVNQKTKTKVLKYEYB---EID  | 380        |              |
| Qy                    | 231              | QANKDIDAACKLATEIA--AIGIKTETETTFYVDYDDILMLSLK                | 275        |              |
| Db                    | 381              | KLKRDLMARDKNGIYLAEEYTGITLKLSONRELNEKMLLKALK                 | 421        |              |

RESULT 10  
US-09-134-001C-5080  
; Sequence 5080, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND  
; TITLE OF INVENTION: EPIMERIDIS FOR D  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,

```

; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

      Query Match      7.2%; Score 108.5; DB 3; Length 3696;
      Best Local Similarity 21.5%; Pred. No. 0.95;
      Matches 71; Conservative 49; Mismatches 128; Indels 83; Gaps 12;

Qy      11 EVKSAETATADGALDLYNKYLDQVIPWKTFDETIKELSR-----FKQEYSQ-EA 58
Db      568 EQNKDIIPSNYTLASNKY-----NKLKERAQTVLDEETNTTFNQFYSTQI 616

Qy      59 SVLVGDIKKVLLM-----DSQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASA 111
Db      617 DLLLHELQTLTLLNKRVSASREINDKAQEMTDVAVDSTELTTEEKDT--LVQDIENHKNEI 673

Qy      112 QKDILRIILDGCVKLLNEAKSKSLTSS-----QSFNASKGLLALDLSQLTNDFS 160
Db      674 SNNIDDELTDGVERVKEAGLHTLESDPHPVTKPNARQVNNRA-----DQOKTLIRN 727

Qy      161 EKSSYFSGQVDRYRK-EAYAGAAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTVQ 219
Db      728 NHEATTEQNEAIRQVEAHSDDAIKIGE-----AETDTTVNEAR 767

Qy      220 NFFTSLSAT-VKQANKIDIAAKLKLATEIAAIGIKETETETTRTFYVDYDDLMLSLKGA 278
Db      768 DNGTKLIATDVPNPTKKAEEA---RAAVTNSANSKIDINNTQATLDERNDALVNRSK 824

Qy      279 KKMINTCNEYQ-----QRHGKKTLEFVP 301
Db      825 DEAIQNTINTAQGNDDVTVEAQNNGNTITQVVP 855

```

```

RESULT 11
US-09-134-001C-4820
; Sequence 4820, Application US/09134001C
; Patent NO. 6380370
; GENERAL INFORMATION:
; APPLICANT: LYNN Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4820
; LENGTH: 1211
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4820

```

|                       |        |  |       |                                    |
|-----------------------|--------|--|-------|------------------------------------|
| Query Match           | 7.1%   | Score 107;   | DB 3; | Length 1211;                       |
| Best Local Similarity | 18.3%; | Pred. No. 0.24;  |       |                                    |
| Matches               | 61;    | Conservative   | 59;   | Mismatches 131; Indels 82; Gaps 9; |
| Qy                    | 7      | EOTVEVVK-----SAIETADGALDLNKYLDVDPWKTFDETIKELSRFKQYQSQASV | 60    |                                    |
|                       |        |  |       |                                    |
| Db                    | 185    | EESNGVLKYKRAESIKLDHDTNLRVVEDILYDLGCRVPLKEBAATAKEVYKQLSKE | 244   |                                    |
|                       |        |  |       |                                    |
| Qy                    | 61     | LVGDIKVLMLDSQDKYFATQTVYENC-----GVVTQLLSAYILLFDSEYN       | 106   |                                    |
|                       |        |  |       |                                    |

Db 245 MEQSDVITVSDIDHYTDNQRLDERLNHLKSOAEKEGQQAQINOLLOKY----- 295  
Qy 107 KKAQAQKDIILRILDDGVKLNKAEQSLTSSQSFNNASGKLALDLSQUTND-----FS 160  
Db 296 -KGRQON-----DYDIEKLN---VELVKATENYEQLSGKLVLEERKKNQSETNARYE 345  
Qy 161 EKSSYFQSDVRIRKEAYAGAAAGIVAGPPGLIISYIAAGVTEGKLIPELNNRL---K 216  
Db 346 EELDNLESQDSIKNEKAQNE-----KLLADLNKKQKQLNK 381  
Qy 217 TVQNFTSLSATVQKANKDIDAAKKLKATETATTAIGETKTTETTRFYVDYDIDLMLSLKKG 276  
Db 382 EVOELSLLYISDQHEKLEBKINSYITLMSQSVVNDIRPLEHTINENEAKSRDLS 441  
Qy 277 AAKMINTCNEYQQ-----RHGKKTLEFV 300  
Db 442 RLVEAFNQLKDIIQNIITQTKYQSSKKMEKV 474

RESULT 12  
US-08-592-126-148  
; Sequence 148, Application US/08592126  
; Patent No. 5821091  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Dolganov  
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
; TITLE OF INVENTION: Polypeptides  
; NUMBER OF SEQUENCES: 151  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,126  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 148:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1312 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54  
US-08-592-126-148

Query Match 7.0%; Score 106.5; DB 2; Length 1312;  
Best Local Similarity 18.9%; Pred. No. 0.3;  
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;  
Qy 10 VEVKSAETADGALDLYNKVLDQVLPKTFDETIK-----ELSRFQKEY 54  
Db 502 MEVLSQNEKAD--LDRTLRKLDQEMQLNHHHTTRTQMEMLTKDKADQDEQIKIKSRH 559  
Qy 55 SQEASVLGDIKVLMSQDKFYFATQTVYVCGVTTQLLSAYILLDFEYNEKKAQKD 114  
Db 560 SDELTSLAG-----YFPNKKQLEDWLHKS-----EINQTR----- 591

Qy 115 ILIRILDDGVKLNKAEQSLTSSQSFNNASGKLALDLSQUTNDFSEK-----SSYFQS 168  
Db 592 -----DRLAKLN---KELASSEQNKHINNELKREEQSS--YEDKLPDVCQSODFES 640  
Qy 169 QVDRIRKE-----AYAGAAAGIVAGPP----- 190  
Db 641 DLDRLEKEIEKSKORAMLAGATA--VYSOFITQLTDENOSCCPVCORVPQTEAELOQEV 698  
Qy 191 -----GLI-ISYSIAAGVIEGKLIPELNNRLKT 217  
Db 699 SLOSKLRLAPDKLKSTESLKKKEKRDDEMLGLVPMRQSIID--LKEKEIPELNNKLON 756  
Qy 218 VQNFTSLSATVQKANKDIDAAKKLKATETATTAIGETKTTETTT-----RFVVDY 266  
Db 757 V-----NRDIQRLKNDIEEQETLLGTIMPEBSAKVCLTDTVIMERFOMEL 802  
Qy 267 DDLMLSLKGAQK-----KMNINTCNEYQORHGKK 295  
Db 803 KDVERKIAQAQKLGQIDLDRTVQOVNQEKQEKQHK 838

RESULT 13  
US-08-687-080-51  
; Sequence 51, Application US/08687080  
; Patent No. 5965427  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Dolganov  
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 175  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/687,080  
; FILING DATE: 17-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/592,126  
; FILING DATE: 26-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0111.30  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1312 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.  
; INDIVIDUAL ISOLATE: 389 TO 4324  
US-08-687-080-51

Query Match 7.0%; Score 106.5; DB 2; Length 1312;  
Best Local Similarity 18.9%; Pred. No. 0.3;  
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;



```

; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18798
; LENGTH: 924
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18798

Query Match      7.0%; Score 105.5; DB 4; Length 924;
Best Local Similarity 17.8%; Pred. No. 0.22;
Matches 51; Conservative 58; Mismatches 100; Indels 77; Gaps 8;

QY 226 SATVQKQNKDIDAAKLKLAIEAIGIKT-----ETETTRFYVDYDDL-----MLSLK 275
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
406 TDAIPK-DRDIHKSNTLPOLYQILKNLRVDRHAMEFEETIETIYMTFDELGGIKELPRTR 464
QY 276 GAAKKMINTC-----NEYQQRHGKKTLEFV 300
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
465 NDAHKLIEECMLLANVAABEVALEHDIPMLYRV 497

RESULT 16
US-09-710-279-1660
; Sequence 1660, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 1660
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-1660

Query Match      7.0%; Score 106; DB 4; Length 885;
Best Local Similarity 18.5%; Pred. No. 0.19;
Matches 54; Conservative 51; Mismatches 111; Indels 76; Gaps 8;

QY 42 ETIKELSRFKQYSQASVILGDIKVLMDSDQKVFATQTVYWC----- 87
Db 2 EPLKEAATAKQYKQSKEMESDVIIVSDIDHTEDNQRLDERNLHLSQQAKEGQ 61
QY 88 GVTYQLLSAYILLDFEYNEKKAQAQKILIRILDGVKKNLPAQKSLTSSQFNASGK 147
Db 62 AQINQLLOKY-----KGRQON-----DYDIEKLN---YELVKATENYEQLSGK 102
QY 148 LLALDSQLTND-----FSEKSSYFQSOVDRIKKEAYAGAAAGIVAGPGLIISYSIAAG 201
Db 103 LNVLEERKKNQSETNARYEEELDNLESQIDSINKEAQN----- 142
QY 202 VIEGKLPIELNNRL---KTVQNFFTSLSATVKQANKDIDAAKLKLAIEAIGIKTET 257
Db 143 -----KLAELKNKQKQLNKEVQELSLIYSDSDHDEKLEIKNSYITLMSQSDVNNDI 198
QY 258 ETTRFYVDYDDLMLSLKGAKKMINTCNEYQQ-----RHGKKTLEFV 300
Db 199 RFLHTTINEAKKSRLDSRLVAFNQLKDIOQNIQTQTOKEYQSSKSMKV 250

RESULT 17
US-09-248-796A-18798
; Sequence 18798, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
```

```

; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18798
; LENGTH: 924
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18798

Query Match      7.0%; Score 105.5; DB 4; Length 924;
Best Local Similarity 17.8%; Pred. No. 0.22;
Matches 51; Conservative 58; Mismatches 100; Indels 77; Gaps 8;

QY 26 LYNKYLDQVTPWKTFFETIKELSR-----FKOYSQASVILGDIKVLMDSDQKYPEATQ 81
Db 233 LHKYVD-----SIRDLKQDFLAFKQEAAGIINVL-NDLDAQSLKQKQIDLEN 281
QY 82 TVYEWCGVVTOLLISAYILLDFEYNEKKAQAQKILIRILDGVKKNLPAQKSLTSSQSF 141
Db 282 TNND-----QIKELVSEHELOIEKISKDLTEKFLVETQLLSKSHSHVQOQFTKELIAESK 338
QY 142 NNA-----SGKLALDSQLTNDFSEKSSYFQSOVDRIKKEAYAGAAAGIVA 187
Db 339 QQVEELEENLKTIIHAKDSARILELTQLSDAAKEKSES DYKLTD----- 382
QY 188 GPFGLIISYSIAAGVIEGKLPIELNNRLKTVQNFFTSLSATVKQANKDIDAAKLKLAIEI 247
Db 383 -----TSEIVNDLKSQIETLKANLKL-----EEREIQNKKLDQVSEL 421
QY 248 AAIIGIKTETETTRFYVDYDDLMLSLKGAKKMINTCNEYQQRHG 293
Db 422 KELKELKVE-----ELSNLLKLQOELHRKEIELNEQLEKLHG 459

RESULT 18
US-09-538-092-807
; Sequence 807, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormatter Version 0.9
; SEQ ID NO 807
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YP070W
US-09-538-092-807

Query Match      6.9%; Score 104.5; DB 4; Length 566;
Best Local Similarity 25.4%; Pred. No. 0.13;
Matches 57; Conservative 29; Mismatches 91; Indels 47; Gaps 11;

QY 15 SAIEATADGALDLYNKY-----LQOVIPTWTFDETI-KELSRFKQ----- 52
Db 6 SVYETLDSMIELFKDYKPGSTILENITRLCQTGLSEFTSELSNELSLRSTASKIIVDV 65
QY 53 EYSQASVILGDIKVLMDSDQK--YFEATQTVYEWCGVVTOLLISAYILLDFEYNEKKA 110
Db 66 DYNKKQD-RIQDVKVLVLSNDFNDFYQNRDGEHKSNIILLNSLTQKYPDLKAFHNILKFL 124
QY 111 AQOKILIRILDGVKKNL-BAQKSLTSSQFNASGKLALDSQLTNDFSEKSSY---- 165
```



```

125 YLDAYSHIESDSTSHNGSSDKLSOASNASFNN-QGKL-----DLFKYFTLSHYIROC 178
166 FQSQVDRIRKEAYAGAAAGIVAGPFGLIISYSAAGVIEGKLIP 209
179 FQDNCCDFKVRTNLNDKFGI-----YILTQG-INGKEVP 211

RESULT 19
US-09-107-532A-3944
; Sequence 3944, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107.532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3944:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...1196
; SEQUENCE DESCRIPTION: SEQ ID NO: 3944:
US-09-107-532A-3944

```

```

Qy      170 VDRIRKAYAGAAAGIVAGPFLGISYSIAAGVIEGKLIPELNRLKTKVQNFTLSATV 229
      |::
      |::
Db      307 KDVLQERT-----KHTQKSSQEQYQTSIAEAQ 332
      |::
      |::
Qy      230 KQAN--KDIDAAKLKL-----TEI--AAIGEIKTETETTRFVVDVDDLMSL 273
      |::
      |::
Db      333 KVKHFPEKQESLMKAAAEKETEIOKAEANLIKTOQELEKYOKSTPKLLAE 384
      |::
      |::

RESULT 20
US-09-540-236-2753
; Sequence 2753, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXY
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2753
; LENGTH: 718
; TYPE: PRT
; ORGANISM: M.catarrhalis
; US-09-540-236-2753

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Rel. #1.0, ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,831C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: PIERRI, MARGARET A.
REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: SIM-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: /desc = "RHAMM I-2a"
US-08-477-831C-11

Query Match      6.8%; Score 103; DB 4; Length 631;
Best Local Similarity 19.8%; Pred. No. 0.22;
Matches 66; Conservative 69; Mismatches 131; Indels 68; Gaps 17

QY      1 MTSIPAGQTVVEVKSATETADGALDLYNKYLDQVDPWKTDFETIKLSRFKOEYSOEASV 60
DB      253 LDNLREKEVEKHIAAQAALIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 309
QY      61 LVGDIKVLMDSDQKYFEATQTVVEGCVVTVQLLSAYIILFDYNEKKKASAKDILIRIL 120
DB      310 L-RDVTAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQSLRDVTAQ-L 359
QY      121 DGVVKLNEAKQSL---LTSQSPNNASGKLALDLSQLTN-DFSEKSSYFOSQVDRIKE 176
DB      360 ESVQEKYNDTAQSLRDVTAQLESYKSSLTKEIE-DUKLENLTQEKVMAEKSVEDVQQQ 418
QY      177 AVAGAAAGIVAGPGLIISYISAGVIEGKLIPELNNR-----LKTQVQNF-----TSL 225
DB      419 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEEIKETSSPLEKITDL 461
QY      226 SATVQKQNKDIDAAKLKLATEIAAIGIKTE--TEITRFVVDYDDL----- 269
DB      462 KNQLRQDQDFRKQLEEKGRKTAENVMTELTWINKWELLYDELYEKTQPFQOQLDAF 521
QY      270 ---MLSLLK--GAAKMINTCNE-YQORHGKKTIL 297
DB      522 EAEKQALLNEHGATQEQQLNKTRDSYAQLLGHQNL 555

RESULT 23
US-09-538-092-1078
; Sequence 1078, Application US/09538092
; Patent No. 675314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqFormatter Version 0.9
; SEQ ID NO 1078

```

```
; LENGTH: 1976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P35580
US-09-538-092-1078

Query Match          6.7%; Score 102; DB 4; Length 1976;
Best Local Similarity 21.1%; Pred. No. 1.6;
Matches 56; Conservative 53; Mismatches 88; Indels 68; Gaps 10;

QY 52 QEYSQASVLVGDIKVLLMDSQKYEATQTVYEWGVVQTLSSAYILLDFEYNEKASA 111
Db 972 EKVTAEAKIKMEEIILLDDQSKTKKEKLM-----DRIACSSQLAESEKAKNLA 1026
QY 112 ----QXDLIRLLDDGVKLLNEAQKSLTSSQSFNNASGKLLALDSQLTNDFSEKSSYFQ 167
Db 1027 KIRKQEVMSLDLEERLKKEKTRQBLEKARK-----LDGE-TTDLQDQIAELQ 1075
QY 168 SQVDRI-----RKEAYAGAAAGIVAGPFGLLIISYSIAAGVIRG-----KLIPELNRR 214
Db 1076 AQIDELKQLAKKEEELQGA-----LARGDDETLHKNNALKVVRLELQAQ 1119
QY 215 LKTVQNPFTSLATVQKANDIDAAKLLKATEIAAAGEIKETETTRFYVDYDDMLSL 274
Db 1120 IABLQEDFSEKASRNKAEQ-----KRLSELEA---LKTELEDT-----LD 1160
QY 275 KGAAKWMINTCNEYQVQHGKKTLP 299
Db 1161 TTAQQELRTKREQVAVELKALEE 1185

RESULT 24
US-08-402-217A-3
; Sequence 3, Application US/08402217A
; Patent No. 5587301
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,217A
; FILING DATE: 10-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PP-0028US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; LENGTH: 1976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P35580
US-09-538-092-1078

Query Match          6.7%; Score 101; DB 1; Length 477;
Best Local Similarity 20.1%; Pred. No. 0.22;
Matches 67; Conservative 72; Mismatches 127; Indels 68; Gaps 18;

QY 1 MTSIPAEQVEVVVKSIAETADGALDLYNKYLDQVFPWKTDFDETIKELSRFKQYSQEASV 60
Db 99 LONLLREKEVELEKHAHAQAAILIAQEKYNTA---QSLRDVTAQLESQEKYKNDTAQS 155
QY 61 LVGDIKVLMDSQDKYEATQTVYEWGVVQTLSSAYILLDFEYNEKKAQAQDILIRIL 120
Db 156 L-RDVTVAQLESQEKYKNDTAQSLRD---VTAQLESQ---EKYND-TAQLSLRDVTAQ-L 205
QY 121 DGQVKLNEAQKSLTSS---QSFNNASGKLLALDSQLTN-DFSEKSSYFQSOVDRIKE 176
Db 206 ESQEKYNDTAQSLRDVSAQLESYKSTLKEIB-DLKJENLTLOEKVAMAEKSVEDVQQQ 264
QY 177 AVAGAAAGIVAGPFGLLIISYSIAAGVIEGKLIIPELNRR-----LKTVOVNF- 225
Db 265 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEEIKETSSFEKITDL 307
QY 226 SATVQKANDI-----DAKLLKATEIA-----AIGEIKTETTRFYVD-Y 266
Db 308 KNQLRQODEDFRKLQBEKGKRTAEKENVMTLMEINKWRLLYEELVEKTKPFQOQLDAF 367
QY 267 DDLMLSLK--GAKKWMINTCNE-YOORHGKKTLP 297
Db 368 EAEKQALLNEHGATQBLNKIRDSYAQLLGHQNL 401

RESULT 25
US-08-700-178-3
; Sequence 3, Application US/08700178
; Patent No. 5783669
; Patent No. 5783669 5700912
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,178
; FILING DATE: August 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/402,217
; FILING DATE: March 10, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0028-1 DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
```

```

; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; LIBRARY: mouse
; CLONE: GI 53979
; US-08-700-178-3

Query Match          6.7%; Score 101; DB 1; Length 477;
Best Local Similarity 20.1%; Pred. No. 0.22;
Matches 67; Conservative 72; Mismatches 127; Indels 68; Gaps 18;

QY 1 MTSIFAEOTVEVVKSAIETADGALDLYNKYLDQVTPKTFDETIKELSRFKQYSQASV 60
DB 99 LDNLLREKEVELEKHIAHAQAIIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKVLMSQDKYFEATQTVYEGCVVTTQLLSAYILLFDEYNEKASAKDILIRIL 120
DB 156 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLSLRDVTAQ-L 205
QY 121 DGVKVLNEAOKSLTSS---QSFNNSAGKLLALDSQLTN-DPSEKSSYFQSOVDRIKE 176
DB 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIE-DLKLENLTLOEKVMAEKSVEDVQQQ 264
QY 177 AVAGAAAGIVAGPFLGIISYIAAGVIEGKLIPELNNR-----LKTQVONFF---TSL 225
DB 265 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEIEKITSFLEKITDL 307
QY 226 SATVQKANKDI-----DAAKKLATEIA-----AIGEIKETEITTFYVD-Y 266
DB 308 KNQLRQODEDFRKQLEEKGRKTAENVMTELMEINKWRLLYBELYEKTKPFQQLDAF 367
QY 267 DDLMLSLK--GAACKWINTCNE-YQORHGKKTIL 297
DB 368 EAEKQALLNEHGATQEQLNKIRDSYAQLLGHQNL 401

```

```

RESULT 26
US-08-995-654-3
; Sequence 3, Application US/08995654
; Patent No. 6025138
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig
; APPLICANT: Seilhamer, Jeffrey
; TITLE OF INVENTION: HVALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,654
; FILING DATE: December 22, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/700,178
; FILING DATE: August 20, 1996
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 08/402,217
; FILING DATE: March 10, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0028-2 DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-995-654-3

Query Match          6.7%; Score 101; DB 3; Length 477;
Best Local Similarity 20.1%; Pred. No. 0.22; 127; Indels 68; Gaps 18;
Matches 67; Conservative 72; Mismatches 127; Indels 68; Gaps 18;

QY 1 MTSIFAEOTVEVVKSAIETADGALDLYNKYLDQVTPKTFDETIKELSRFKQYSQASV 60
DB 99 LDNLLREKEVELEKHIAHAQAIIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKVLMSQDKYFEATQTVYEGCVVTTQLLSAYILLFDEYNEKASAKDILIRIL 120
DB 156 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLSLRDVTAQ-L 205
QY 121 DGVKVLNEAOKSLTSS---QSFNNSAGKLLALDSQLTN-DPSEKSSYFQSOVDRIKE 176
DB 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIE-DLKLENLTLOEKVMAEKSVEDVQQQ 264
QY 177 AVAGAAAGIVAGPFLGIISYIAAGVIEGKLIPELNNR-----LKTQVONFF---TSL 225
DB 265 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEIEKITSFLEKITDL 307
QY 226 SATVQKANKDI-----DAAKKLATEIA-----AIGEIKETEITTFYVD-Y 266
DB 308 KNQLRQODEDFRKQLEEKGRKTAENVMTELMEINKWRLLYBELYEKTKPFQQLDAF 367
QY 267 DDLMLSLK--GAACKWINTCNE-YQORHGKKTIL 297
DB 368 EAEKQALLNEHGATQEQLNKIRDSYAQLLGHQNL 401

```

```

RESULT 27
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; TITLE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-310-187A-1

Query Match          6.7%; Score 101; DB 3; Length 1939;
Best Local Similarity 22.5%; Pred. No. 1.9; 128; Indels 86; Gaps 13;
Matches 75; Conservative 44; Mismatches 128; Indels 86; Gaps 13;

QY 5 FAEQTVVEVVKSAIETADGALDLYNKYLDQVTPKTFDETIKELSRFKQYSQASV 64

```

Db 1173 FQMRDLLEATLQHEATAAALRKKHADV---AELGEQIDNLRVQKLEKEKS-----E 1225  
Qy 65 IKVLLMDSQDKYFEATQTVYVCGVTTQLLSAYILLFDEYNEKKASAKQDILIRILDGV 124  
Db 1226 FKLELDD-----VTSNMEQII-----KAKANLEKVSRTLEDQA 1258  
Qy 125 K-----KLINEAKSL---LTSSQSFNNASGKLL-----ALDSQLTNDPSEKSSYFOSQV 170  
Db 1259 NEYRVKLEEAQRSINDFTTQRAKLQTEGELARQLEKEKALISQLTRG---KLSYQQOME 1315  
Qy 171 DRIRKEAYAGAAAGIVAGPGLIISYISIAAGVIEGKLIPE-----LNRL 215  
Db 1316 DLKRLQEEGKAKNALA-----HALQSAHDCDLLREQYEEETEAKAELQVLSKAN 1367  
Qy 216 KTVQNFTSLSATVKQANKDIDAUKLATE-----IAAIGIKETETETT---RFVVDY 266  
Db 1368 SEVAQWRTKYETDAIQRTTELEEAQKKLAQRLQDAEBAVAVNAKCSLSLEKTKHRLQNEI 1427  
Qy 267 DDLMLSLKGAACKMINTCNEYQORHGKKTLPF 299  
Db 1428 EDLMVDVERSNA---AALDKQRNFDKILAE 1457  
RESULT 28  
US-09-538-092-917  
; Sequence 917, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormatter Version 0.9  
; SEQ ID NO 917  
; LENGTH: 1939  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P13533  
US-09-538-092-917

Query Match 6.7%; Score 101; DB 4; Length 1939;  
Best Local Similarity 22.5%; Pred. No. 1.9;  
Matches 75; Conservative 44; Mismatches 128; Indels 86; Gaps 13;  
Qy 5 FAEQTVVVKSAIETADGALDLYNKYLDQVLPKTFDETIKELSRFKQYEQEASVLGD 64  
Db 1173 FQMRDLLEATLQHEATAAALRKKHADV---AELGEQIDNLRVQKLEKEKS-----E 1225  
Qy 65 IKVLLMDSQDKYFEATQTVYVCGVTTQLLSAYILLFDEYNEKKASAKQDILIRILDGV 124  
Db 1226 FKLELDD-----VTSNMEQII-----KAKANLEKVSRTLEDQA 1258  
Qy 125 K-----KLINEAKSL---LTSSQSFNNASGKLL-----ALDSQLTNDPSEKSSYFOSQV 170  
Db 1259 NEYRVKLEEAQRSINDFTTQRAKLQTEGELARQLEKEKALISQLTRG---KLSYQQOME 1315  
Qy 171 DRIRKEAYAGAAAGIVAGPGLIISYISIAAGVIEGKLIPE-----LNRL 215  
Db 1316 DLKRLQEEGKAKNALA-----HALQSAHDCDLLREQYEEETEAKAELQVLSKAN 1367  
Qy 216 KTVQNFTSLSATVKQANKDIDAUKLATE-----IAAIGIKETETETT---RFVVDY 266  
Db 1368 SEVAQWRTKYETDAIQRTTELEEAQKKLAQRLQDAEBAVAVNAKCSLSLEKTKHRLQNEI 1427

Qy 267 DDLMLSLKGAACKMINTCNEYQORHGKKTLPF 299  
Db 1428 EDLMVDVERSNA---AALDKQRNFDKILAE 1457  
RESULT 29  
US-09-538-092-844  
; Sequence 844, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormatter Version 0.9  
; SEQ ID NO 844  
; LENGTH: 643  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P04264  
US-09-538-092-844

Query Match 6.6%; Score 100.5; DB 4; Length 643;  
Best Local Similarity 22.2%; Pred. No. 0.4;  
Matches 72; Conservative 58; Mismatches 124; Indels 71; Gaps 15;  
Qy 5 FAEQTVVVKSAIETADGALDLYNKYLDQVLPKTFDETIKELSRFKQYEQEASVL 61  
Db 199 FLEQGNQVLQTKWELLQQVDTSRTHN-----LEPY---FESFINLRRRVYDQLKSDQSL 251  
Qy 62 VGDIK---VLLMDSQDKYPE---ATQTVYVCGVTTQLLSAYILLFDEYNEKKASAKQD 114  
Db 252 DSELKQMDVEDYRNKYDEINKRTNAENEFYTIKVDGATYTKVD-LQAKLDNLQOE 310  
Qy 115 ILIRILDGVKKLNEAKSLTSSQSFNNA-----SGKLLALDSQL-----TNDFSEKS 163  
Db 311 I-----DPLTALYQAELSQMOTQISETNVILSMNDRSLDLSIIAEVKAQNEIDIAQS 364  
Qy 164 -----SYFQSQVDRIKREYAGAAAGIVAGPFG-----LIISYISIAAGVIEGKLIPELNN 213  
Db 365 KAAEASLYQSKYBELQ-----ITAGRHGDSVRNSKIEISELNRVIQ-RLRSIDN 413  
Qy 214 RLKTVQNFTSLSATVKQANKDIDAUKLATEIAAIGETIKETETTRFRVYVDDLM--- 270  
Db 414 VKQISNLOOSISDASQGENALKDKNLDLEALQQAQK--EDLARLLRDYQELMNTK 471  
Qy 271 -----LSLLKGAACKMINTC 285  
Db 472 LALDLEIATVYRTLLEGEBSRMSGEC 496

RESULT 30  
US-09-328-352-5793  
; Sequence 5793, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: BAUMNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04

1166 EQBNILKKTLEBEAKTHEAQIQEMRQKHSQAV--BELAQLEQETRYKVKANLEKAKQTLE 1223  
56 QEASVLVGDIKVLLMDSQDKYFEATQVYEWGVVTLSSAYILLFDEYNEKKAQAQDI 115  
1224 NERGELANEVKKVLLQKGD-----SEHKRKKVEAQLQE 1256  
116 LIRILDDGVKCLNEAOKSLTSSQSNFNASGKLLALD---SOLTNDPSEKSSYFQSOVD 172  
1257 LQVKNFGERVRELADKVTKLQVELDNTGLSSQSDSKSLTKDFFSALESQLDQTOEL 1316  
173 IRKEAYAGAAGIVAGPFGILLISYIAAGVIEGKLIPELNNRLKTVQNFFTLSATVKA 232  
1317 LOEN-----RQKL--SLSTKLQVEDEKNSFREOLEEE 1348  
233 NKDIDAARKLATEIAAIGIKETETTRFYVDYDDI--MLSLKGAAGKMWNTCNEYQOR 291  
1349 EAKHNLEKQIATLHAQVADMKKME-----DSVGCLETAEEVKRKLQKLEGLSQR 1400  
292 HGKK 295  
1401 HBEK 1404  
RESULT 32  
US-09-538-092-936  
; Sequence 936, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormatter Version 0.9  
; SEQ ID NO 936  
; LENGTH: 2871  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P15924  
US-09-538-092-936  
Query Match 6.6%; Score 100; DB 4; Length 2871;  
Best Local Similarity 17.8%; Pred. No. 4.5; Mismatches 57; Indels 128; Gaps 12;  
Matches 60; Conservative 57;  
QY 25 DLYNKYLDQVPMKTFDETIKLSRFKQESQASVVLG-----DIKVLMDSDQKYP 77  
DB 1205 NLRNYETEINIKT-----TIKESMQKEDSKLRLQDLRSRENRDLKDEIVRLNDSIL 1261  
QY 78 EATQ-----TVYEWCGVVTQLLSAYILLFDEYNEKKAQAQ---KDLIRILDDGVK 125  
DB 1262 QATEQRRAENALQKACG-----SEIMQKQHLIELKQWQORSEDNAR 1308  
QY 126 ---KLINEAOKSLTSSQSF-----NNASGKLLALDS--- 153  
DB 1309 HKQSLEAAKTIQDNKEIEELKAEFOEAKRWEYENELSKVNNYDEBIISLKNQFET 1368  
QY 154 -----QLTNDPSEKSSYFQSOVDRIKKEAYAGAAGIVAGPFGILLISYIAAGVI 203  
DB 1369 EINITKTIHQITWQKEEDTSGYRAQIDNLTRENS----- 1404  
QY 204 EGKLPPELNNRLKTVQNFFTLSATVKAQKIDAAKL-----KLATEIAAIGEI 253  
DB 1405 -----LSEETIKRLKNTLTOTENLRARVEEDIQOQKATGSEVRSQKQOLEVELRQVTQM 1457

NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5793  
; LENGTH: 1454  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5793  
Query Match 6.6%; Score 100.5; DB 4; Length 1454;  
Best Local Similarity 19.6%; Pred. No. 1.4; Mismatches 133; Indels 139; Gaps 18;  
Matches 80; Conservative 57;  
QY 11 EVKSAJET-----ADGALDLYNKYLDQVPMKTFDETIKE-----LSRF 50  
DB 278 QVLGAGSVOTLASNLDIADGALVVGIGYITRAILMKS--AAIKEGNASTLASRQASVINA 335  
QY 51 KQESYQSEASVL-----VGDIKVLLMDSQDKYFEAT-----QTVYE 85  
DB 336 QAEYAEATAALNAKAHLANVRATNAETQAK--FGATAAATRYAQAQAAVTAATNAQTAQA 394  
QY 86 -----WGVVVTQLLSAYILLFDEYNEKKAQAQDI-----LI 117  
DB 395 IKLNTATSIAKLAGAFGLICGWAGVATLGVMLAAAYSYFNNAKAEAKQKLAEOAKVA 454  
QY 118 RILDDGVKLL--NEAOKSLTSSQSNFNASGKLLALDPSOLTNDP---SEKSSYFQSOVD 172  
DB 455 EKADBEELKLTGNDKAKAVNDLITAF--NAQNKALEKSRVAGSALIDIENYARGNREVEK 513  
QY 173 IRKEAYAGAAGIVAGPFGILLISYIAAGVIEGKLIPELNNRLKTVQNFFTLSATVKAQ 231  
DB 514 ISQEARTGT-----ISYTEAI-----ERLNKIKLPTDLYENLKQQAQY 552  
QY 232 ---ANKDIDAARKL-----ATEIAAIGIKETETE--TTRFYVDY- 266  
DB 553 DDNASKASLSASBKLLRVEVKLGNEAQAQNAIQHOKQADALGNATATEAKATKALQDYQ 612  
QY 267 -----DDLMLS--LLKGAAGKMWNTCNEYQORHGKTKTFEVPDVNAS 305  
DB 613 AKQKQSVDSIYKSGWLDKGYTVAQANAILELQKAGMSAILSKDEIDS 661  
RESULT 31  
US-09-538-092-1077  
; Sequence 1077, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormatter Version 0.9  
; SEQ ID NO 1077  
; LENGTH: 1960  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P35579  
US-09-538-092-1077  
Query Match 6.6%; Score 100; DB 4; Length 1960;  
Best Local Similarity 19.1%; Pred. No. 2.5; Mismatches 120; Indels 80; Gaps 9;  
Matches 58; Conservative 46;  
QY 7 EQTVKVSIAIE-----TADGALDLYNKYLDQVPMKTFDETIKE-----LSRFKQES 55



APPLICATION NUMBER: US/08/938,105

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Crook, Wamell M.

REGISTRATION NUMBER: 31,071

REFERENCE/DOCKET NUMBER: 3595-4

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1886 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-938-105-3

Query Match 6.5%; Score 98; DB 3; Length 1886;

Best Local Similarity 21.3%; Pred. No. 3.7;

Matches 71; Conservative 44; Mismatches 131; Indels 88; Gaps 12;

5 PAQVTEVVKSAIETADGALDLYNKYLDQVTPWKTFTETIKELSRFKQESQVSLVGD 64

1120 FQKQERDLEATLQHEATAAALRKHADSV---AELGEQIDNLQKVQKLEKES---E 1172

65 IKVLLMSQDKYFEATQVYEWCGVVTQLLSAYILLFDEYNEKKASAKQKILIRILDGV 124

1173 FKLELDD-----VTSHMQII-----KAKANLEKVSRTLEDOA 1205

125 K---KLNEAQSLLTSSQSFNNASGKLLALDSQTNDFSEKS-----SYFSQ 169

1206 NEYRVKLEAQSLL---NDFTQRAKLQTEGELARQLEKEALLIWLTRGKLSYTOOM 1261

170 VDIRKEAYAGAAAGIVAGPFGLLISYSIAAGVIEGKLPE-----LNNR 214

1262 EDLKRQLEBEGKAKNLA-----HALQSAHDCDLREQYEEEMAEKALQRLVLSKA 1313

215 LKTVQNFETSLATVKQANKIDAAKILATE-----IAAIGIKTETET---RFYVD 265

1314 NSEVAGRWTKYETDALQRTTEELAEAKKLAQRLDAEEAVEAVNAKCSLEKTKHRLQNE 1373

266 YDPLMLSLKGAAGKAMINTCNEYQOHGKKTIFE 299

1374 IEDLMVDVERNSAAA---AALDKQORNFDKILAE 1404

US-09-252-991A-31794

Sequence 31794, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 31794

LENGTH: 669

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31794

Query Match 6.4%; Score 97.5; DB 4; Length 669;

Best Local Similarity 21.8%; Pred. No. 0.84;

Matches 57; Conservative 48; Mismatches 94; Indels 63; Gaps 10;

193 VE---SQADGMVAG-----MTITDRKKAPDFSVFVDSGIQIA 229

229 VKQAN---KDIDAAK---LKLATEIA-AIGBIKTETETTRFYVDYDDMLSL 274

230 VKGNDKIKSYDDLKGVKGVKIGTESADPLEKXKKYDYSIKYLTDTTALYSAL 284

RESULT 36

US-09-328-352-5599

Sequence 5599, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 5599

LENGTH: 733

TYPE: PRT

ORGANISM: Acinetobacter baumannii

US-09-328-352-5599

Query Match 6.5%; Score 98; DB 4; Length 733;

Best Local Similarity 22.0%; Pred. No. 0.86;

Matches 65; Conservative 43; Mismatches 86; Indels 102; Gaps 14;

12 VVKSIAETADGALDLYNKYLDQVTPWKTFTETIKELS-----RFKQEYS-QEASVLVGD 64

201 VMKVAIFTKD---QLDSKY-----NKKLSIPAADVNNISSYVAERGKLTG- 244

65 IKVLLMSQDKYFEATQVYEWCGVVTQLLSAYILLFDEYNEKKASAKQKILIRILDGV 124

245 ILGLNVOGYDKEH-----ITQVLNAILVTYGAQNVRRSAESAOTLKFLEQL 292

125 ----KLNEAQSLLTSSQSFNNASGKLLALDSQTNDFSKSSYFQSQVDRI- 174

293 PDLKQLDAAERQFNKFRQYN-----TVDVTKESLYLTQSITLETAKAEL 339

175 KEAYAGAAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTQNFETSLATVKQANK 234

340 EQKQAEWAAYTA-----EHPAMREINGQITAINKQIGELNSTLKQL-P 382

235 DIDAAKILATEIAAIGBIKTETETTRFYVDYDDMLSLKGAAGKAMINTCNEYQQ 290

383 DVQRQYLQLYREV-----EVRKQ-----LYTALL-----NSYQQ 411

RESULT 37

US-08-938-105-3

Sequence 3, Application US/08938105

Patent No. 6353151

GENERAL INFORMATION:

APPLICANT: Leinwand, Leslie A.

APPLICANT: Vikstrom, Karen L.

TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross P.C.

STREET: 1700 Lincoln St., Suite 3500

CITY: Denver

STATE: CO

COUNTRY: U.S.A.

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:



QY 33 QVTPKTFDETIKELSRFQKEY-----SQEASVL-----VGDIKVLLMDSQDK 75  
Db 112 EVVPDKQVDDIDAEIAFLKQALGVRADMRALSSKLASQURKEARALFDVILMLDDASI 171  
QY 76 YFEATQTVY--EWC-GVVTQLLSAYI-----LLFDEYNEKKASQAQKQILIRILDDGVKKLN 128  
Db 172 GNEVKRIIRTGWAQAGALQVVMHVQRFELMDDAYLRRASDVKDIGRELL-----AYLQ 227  
QY 129 EAKSLLT-----SQSFNNA-----SGKLLALDSQLTNDPSEKSSYFQSQVDRIRK 175  
Db 228 EERKQNLTYPEQTIIVSEELSPAMLGEVPEGRVLGLVSLVSG-----NSHVAILAR 279  
QY 176 EAYAGAAAGVAGPFGGLIISYSIAAGVIEGKLIPELNNRLKTVQNPFISATVKQANKD 235  
Db 280 AMGIPTVMGAVDLPYSKVDGIDLIVDGYHGEV-----YTNPSAELVRQYSD 325  
QY 236 IDAAKLKLAIEAIAIGIKTET 257  
Db 326 VVAERELSKGLAALRLPCET 347

## RESULT 39

US-09-199-637A-170  
; Sequence 170, Application US/09199637A  
; Patent No. 6355411

## GENERAL INFORMATION:

; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John

## TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID

## SEQUENCES AND USES THEREOF

; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A

; CURRENT FILING DATE: 1998-11-25

; PRIOR APPLICATION NUMBER: 60/066,517

; PRIOR FILING DATE: 1997-11-25

; NUMBER OF SEQ ID NOS: 437

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 170

; TYPE: PR1

; ORGANISM: Pseudomonas aeruginosa

US-09-199-637A-170

Query Match 6.4%; Score 97.5; DB 3; Length 759;

Best Local Similarity 21.8%; Pred. No. 1;

Matches 57; Conservative 48; Mismatches 94; Indels 63; Gaps 10;

QY 33 QVTPKTFDETIKELSRFQKEY-----SQEASVL-----VGDIKVLLMDSQDK 75  
Db 202 EVVPDKQVDDIDAEIAFLKQALGVRADMRALSSKLASQURKEARALFDVILMLDDASI 261  
QY 76 YFEATQTVY--EWC-GVVTQLLSAYI-----LLFDEYNEKKASQAQKQILIRILDDGVKKLN 128  
Db 262 GNEVKRIIRTGWAQAGALQVVMHVQRFELMDDAYLRRASDVKDIGRELL-----AYLQ 317  
QY 129 EAKSLLT-----SQSFNNA-----SGKLLALDSQLTNDPSEKSSYFQSQVDRIRK 175  
Db 318 EERKQNLTYPEQTIIVSEELSPAMLGEVPEGRVLGLVSLVSG-----NSHVAILAR 369  
QY 176 EAYAGAAAGVAGPFGGLIISYSIAAGVIEGKLIPELNNRLKTVQNPFISATVKQANKD 235  
Db 370 AMGIPTVMGAVDLPYSKVDGIDLIVDGYHGEV-----YTNPSAELVRQYSD 415  
QY 236 IDAAKLKLAIEAIAIGIKTET 257  
Db 416 VVAERELSKGLAALRLPCET 437

## RESULT 40

US-09-248-796A-16013  
; Sequence 16013, Application US/09248796A  
; Patent No. 6747137

## GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 16013

; LENGTH: 862

; TYPE: PR1

; ORGANISM: Candida albicans

US-09-248-796A-16013

Query Match 6.4%; Score 97.5; DB 4; Length 862;

Best Local Similarity 21.8%; Pred. No. 1.2;

Matches 65; Conservative 59; Mismatches 97; Indels 77; Gaps 16;

QY 13 VKSAIETAGALDLYNKYLDQVTPKTFDETIKELSRFQKEYSQEASVLVDGDKVLLMDS 72  
Db 239 VKSAI--AEQAILAFPKNPEVL-----ETIKIDESYWKNLATE-----KAFLM-- 280  
QY 73 QDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASQAQKQILIRILDDGVKKLNEAQK 132  
Db 281 -----RTFYQYCN-ENQLHALMDANPELDLSITLEKYSVRL-----KTINE-NE 325  
QY 133 SILTSSQSFNNAAGKLLALDSQLTNDPSEKSSYFQSQVDRIRKEA--YAGAAAGIVAGPF 190  
Db 326 NLVKTWETYN---AKIDELDNQI---FS-----LENQISRINTDADNFRKLSLSNIED-- 372  
QY 191 GLIISYISIAAGVIEGKLIPELNNRLKTVQNFFT-----SLSATVTKQANK 234  
Db 373 --IIEINIAKDLFK-KRIKQKNNSGNLEDLITEENQEADQIKDFLMEDLQQOLEDINK 429  
QY 235 DIDAAKLKLATEIAAIGEIKTETETTRFYVDYDDLLMLLKGAAKMKINTCNEVQQRH 292  
Db 430 NLDEIEHQPEDITAKLEELQTK-----YDS-CIRALETTSELKIQTQVIFEOAH 477

## RESULT 41

US-09-917-254-91

; Sequence 91, Application US/09917254

; Patent No. 6703204

## GENERAL INFORMATION:

; APPLICANT: Mutter, George  
; APPLICANT: Baak, Jan  
; TITLE OF INVENTION: Prognostic Classification of Breast Cancer  
; FILE REFERENCE: B0801/7224 (JRV)  
; CURRENT APPLICATION NUMBER: US/09/917,254  
; CURRENT FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 60/222,093  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 91

; LENGTH: 1857

; TYPE: PR1

; ORGANISM: Homo Sapiens

US-09-917-254-91

Query Match 6.4%; Score 97.5; DB 4; Length 1857;

Best Local Similarity 21.0%; Pred. No. 4.1;

Matches 66; Conservative 51; Mismatches 108; Indels 89; Gaps 13;

7 EQTEVVVKSATADGA-----LDLYNKYLDQVPIPKTFDETIKELSRF-----K 51  
1058 EQTEVVVKSATADGA-----LDLYNKYLDQVPIPKTFDETIKELSRF-----K 51  
52 QEQVTVLKKALDETRSHQAQVQEMRQKHAQV-----EELTEQLEQFKRAKANLDK 1111  
1112 QTEKENADLAGELRVL-----GQAKQEV-----EELTEQLEQFKRAKANLDK 1111  
112 QKDLIRILDGVK-----KLNEAQSKLLTSSQFNNASGKLALDLSQTLNDFSEK 164  
1145 QVQELQSKCSGGERARAEKNDKVKHLQNEVESVTGMLNEAGKAI-----KLAKDVASLSS 1200  
165 YFQSQVDRIKAEVAGAAAGIVAGPFGLLIISYSIAGVIE---GKLIPELNNELKTQVNF 221  
1201 QLQDTQELLQETTRQK-----LNVTSLRQLREERNLQDQDEMEAKQN- 1246  
222 FTSLSATVQKANKDIDAUKL-----ATEIAAIGBIKTETTRFYVDYDDMLSLKGA 278  
1247 ---LERHISTLNIQLSDSKKKLQDFASTVLEAEGK-----KRFQKEIENLTQOYEKAA 1298  
279 --KEMINTCNEYQ 290  
1299 AYDKLEKTKNRLQ 1312

RESULT 42  
US-09-538-092-1084  
; Sequence 1084, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Glot, Loic  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuratSeqFormat Version 0.9  
; SEQ ID NO 1084  
; LENGTH: 1972  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P35749  
US-09-538-092-1084

Query Match 6.4%; Score 97.5; DB 4; Length 1972;  
Best Local Similarity 21.0%; Pred. No. 4.4; 108; Indels 89; Gaps 13;  
Matches 66; Conservative 51; Mismatches 51; Mismatches 108; Indels 89; Gaps 13;  
7 EQTEVVVKSATADGA-----LDLYNKYLDQVPIPKTFDETIKELSRF-----K 51  
1173 EQTEVVVKSATADGA-----LDLYNKYLDQVPIPKTFDETIKELSRF-----K 51  
52 QEQVTVLKKALDETRSHQAQVQEMRQKHAQV-----EELTEQLEQFKRAKANLDK 1226  
1227 QTEKENADLAGELRVL-----GQAKQEV-----EELTEQLEQFKRAKANLDK 1226  
112 QKDLIRILDGVK-----KLNEAQSKLLTSSQFNNASGKLALDLSQTLNDFSEK 164  
1260 QVQELQSKCSGGERARAEKNDKVKHLQNEVESVTGMLNEAGKAI-----KLAKDVASLSS 1315  
165 YFQSQVDRIKAEVAGAAAGIVAGPFGLLIISYSIAGVIE---GKLIPELNNELKTQVNF 221  
1316 QLQDTQELLQETTRQK-----LNVTSLRQLREERNLQDQDEMEAKQN- 1361  
222 FTSLSATVQKANKDIDAUKL-----ATEIAAIGBIKTETTRFYVDYDDMLSLKGA 278

1362 ---LERHISTLNIQLSDSKKKLQDFASTVLEAEGK-----KRFQKEIENLTQOYEKAA 1413  
279 --KEMINTCNEYQ 290  
1414 AYDKLEKTKNRLQ 1427  
RESULT 43  
US-09-308-375-2  
; Sequence 2, Application US/09308375  
; Patent No. 6300117  
; GENERAL INFORMATION:  
; APPLICANT: Genencor International, Inc.  
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms  
; FILE REFERENCE: GC394-PCT  
; CURRENT APPLICATION NUMBER: US/09/308,375  
; CURRENT FILING DATE: 1999-05-14  
; EARLIER APPLICATION NUMBER: EP9719636.4  
; EARLIER FILING DATE: 1997-09-15  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 2285  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-308-375-2

Query Match 6.4%; Score 97.5; DB 3; Length 2285;  
Best Local Similarity 19.9%; Pred. No. 5.6;  
Matches 65; Conservative 58; Mismatches 106; Indels 97; Gaps 16;  
9 TVEVVKSAIETADGAL-----DLYNKYLDQVPIPKTFDETIKELSRFQKQESQ 57  
1091 SVDDIKAAIKEMSDAMQFDSVDVLNGDIFNNTKQVAP---LNDLLEKMAEGKISANE 1147  
58 ASVLVGDIVKLLMDSQDKYFEATQTVYEWCGVV---TQLLSAYILLFDEYNE---K 107  
1148 ANTLI-----QKDK--ELAQAIENGWVVKINRDEVIKQKVKLDAYNDVMVYSNK 1196  
108 KASAKDILIRILD-----DGVKKLNEAQSKLLTSSQ-----SFNNASGKLALD 154  
1197 LMKTEVNNAIKTLNADTLRLIDSLKLRKIDMSEALSDLEVKSIINNVADAKKEL-KK 1255  
155 LTNDFSEKSSYFQSOVDRIK-----EAYAGAAAGIVAGPFGLLIISYSIAGVIEGKLIP 209  
1256 LEEKMLQPGYSNSQIEAMQSVKSALESYISASE-----EATSTQ 1295  
210 ELNNRLKTQNFFTSLSATVQKANKDIDAUKLATEIAAIGBIKTETTRFYVD-YDD 268  
1296 EMNK--QALVEAGTSLNWTDDQEK-----ANE-----ETKTSMTVVVDKYKE 1335  
269 LMLSLKGAAG--KEMINTCNEYQORH 292  
1336 ALEKVNABIDKYNKQVNDYPKYSQY 1361

RESULT 44  
US-08-328-254-6  
; Sequence 6, Application US/08328254  
; Patent No. 5710022  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Xueliang  
; APPLICANT: Lee, Wen-Hwa  
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-254-6

Query Match      6.4%; Score 97.5; DB 1; Length 2482;
Best Local Similarity 20.5%; Pred. No. 6.3;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;

QY  7 EQTVVVVKSIAETADGALDYN----KYLDQVIPW-----KTFDETIKELSRFKQY 54
Db  1948 QDTLEVLSQSSYKLNLELELTMDKMSFVEKVNKMTAKETELQREHMAQKTAELQEE 2007
QY  55 SQEASVLVGDIVKVLMD--SODKYFEAT-----QTVVEMGVVTVLLSAY 97
Db  2008 SGKNRLAGELQLLLEIKSSQDKQLKELTLENSELKSLDCMHKQDQVEKGGKVEETAEY 2067
QY  98 ILLFDEYNEKKASA-----QKDILIRILDGCVKKNLNEAQS-----LLTSSOSFNN 143
Db  2068 QLRLHE-AEKHQALLDNTNKQVEVEIQTYREKLTSSQKLEIDLLKSSKEELNN 2126
QY  144 ASGKLALDSQLTNDFSEKSSYFQSQVDRIKRAYAGAAAGIVAGPGLIISYSIAAGVI 203
Db  2127 S-----LKATTTQILEELKTKMDNLKYNQLKENER-----AQGKMKLLI----- 2167
QY  204 EGKLIPELNNRLKTVQNFPTSL-SATVQKANKDIDAACL-KLATEIAAIGS-IKTETETT 260
Db  2168 --KSCQLEBEKEITQKELSQLQAQEQKGTGTVMDTKVDELTEIKELKETLEBKTEA 2225
QY  261 RFYVD-YDDLMLSLK-GAAKGMINT-----CNEYQQRHGK 294
Db  2226 DEYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQSQSQKQDSR 2266

RESULT 45
US-09-538-092-1154
; Sequence 1154, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
```

```
; SEQ ID NO 1154
; LENGTH: 3210
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)_(0)
; OTHER INFORMATION: Polypeptide Accession Number P49454
; US-09-538-092-1154

Query Match      6.4%; Score 97.5; DB 4; Length 3210;
Best Local Similarity 20.5%; Pred. No. 9.4;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;

QY  7 EQTVVVVKSIAETADGALDYN----KYLDQVIPW-----KTFDETIKELSRFKQY 54
Db  2676 QDTLEVLSQSSYKLNLELELTMDKMSFVEKVNKMTAKETELQREHMAQKTAELQEE 2735
QY  55 SQEASVLVGDIVKVLMD--SODKYFEAT-----QTVVEMGVVTVLLSAY 97
Db  2736 SGKNRLAGELQLLLEIKSSQDKQLKELTLENSELKSLDCMHKQDQVEKGGKVEETAEY 2795
QY  98 ILLFDEYNEKKASA-----QKDILIRILDGCVKKNLNEAQS-----LLTSSOSFNN 143
Db  2796 QLRLHE-AEKHQALLDNTNKQVEVEIQTYREKLTSSQKLEIDLLKSSKEELNN 2854
QY  144 ASGKLALDSQLTNDFSEKSSYFQSQVDRIKRAYAGAAAGIVAGPGLIISYSIAAGVI 203
Db  2855 S-----LKATTTQILEELKTKMDNLKYNQLKENER-----AQGKMKLLI----- 2895
QY  204 EGKLIPELNNRLKTVQNFPTSL-SATVQKANKDIDAACL-KLATEIAAIGS-IKTETETT 260
Db  2896 --KSCQLEBEKEITQKELSQLQAQEQKGTGTVMDTKVDELTEIKELKETLEBKTEA 2953
QY  261 RFYVD-YDDLMLSLK-GAAKGMINT-----CNEYQQRHGK 294
Db  2954 DEYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQSQSQKQDSR 2994

RESULT 46
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
```

us-09-993-292b-2.ra1

Wed Jan 5 14:16:58 2005

```
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORGANISM: HUMAN
PCT-US95-16216-1

Query Match 6.4%; Score 97.5; DB 1; Length 3248;
Best Local Similarity 20.5%; Pred. No. 9.6;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;

QY 7 EQTVVVKSAIETADGALDLYN----KYLDQVIPW-----KTFDETTELKSRPKQEV 54
Db 2676 QDTLEVLQSSYKNLENELELTMDKMSFVEKVNQNTAKETELQREHMAQKTAELQEL 2735

QY 55 SOEASVLVDIKVLLMD---SODKYFEAT-----QTVVWCGVVTQLLSAY 97
Db 2736 SGEKNRLAGELQLLLEIKSKQQLKELTLENSELKKSLLDCMHKQVKEGKVRERIAEY 2795

QY 98 ILLDFEYNEKASA-----QKDLIRILDGKVKLINEAKS-----LLTSSQSFNN 143
Db 2796 QLRLHE-AEKHQALLDNTNKQYEVEIQTYREKLTSEKLSOKLEIDLKSKKEELNN 2854

QY 144 ASGKLALDSQLTNDFSEKSSVFQSOVDRIKRAYAGAAAGIVAGPFGLIISYIAAGVI 203
Db 2855 S-----LKATTLLEBKTKMDNLKYVNLKENER-----AQGMKLLI----- 2895

QY 204 EGKLIPELNNRLKTVQNFPTSL-SATVQANKDDAKL-KLATEIAAIGE-IKTETETT 260
Db 2896 --KSCQLEEEKEILQKELSQLQAAQEKQKTGTVMDTKVDLTTEIKELKETEETKEA 2953

QY 261 RFYVD-YDDLMLSLK-GRAKWMINT-----CNEYQORHGK 294
Db 2954 DEYLDKYCLLSISHEKLEKAKEMLETTQVNLCSQSQSKDSR 2994

RESULT 47
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1

Query Match 6.4%; Score 97.5; DB 5; Length 3248;
Best Local Similarity 20.5%; Pred. No. 9.6;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;

QY 7 EQTVVVKSAIETADGALDLYN----KYLDQVIPW-----KTFDETTELKSRPKQEV 54
Db 2676 QDTLEVLQSSYKNLENELELTMDKMSFVEKVNQNTAKETELQREHMAQKTAELQEL 2735

QY 55 SOEASVLVDIKVLLMD---SODKYFEAT-----QTVVWCGVVTQLLSAY 97
Db 2736 SGEKNRLAGELQLLLEIKSKQQLKELTLENSELKKSLLDCMHKQVKEGKVRERIAEY 2795

QY 98 ILLDFEYNEKASA-----QKDLIRILDGKVKLINEAKS-----LLTSSQSFNN 143
Db 2796 QLRLHE-AEKHQALLDNTNKQYEVEIQTYREKLTSEKLSOKLEIDLKSKKEELNN 2854

QY 144 ASGKLALDSQLTNDFSEKSSVFQSOVDRIKRAYAGAAAGIVAGPFGLIISYIAAGVI 203
Db 2855 S-----LKATTLLEBKTKMDNLKYVNLKENER-----AQGMKLLI----- 2895

QY 204 EGKLIPELNNRLKTVQNFPTSL-SATVQANKDDAKL-KLATEIAAIGE-IKTETETT 260
Db 2896 --KSCQLEEEKEILQKELSQLQAAQEKQKTGTVMDTKVDLTTEIKELKETEETKEA 2953

QY 261 RFYVD-YDDLMLSLK-GRAKWMINT-----CNEYQORHGK 294
Db 2954 DEYLDKYCLLSISHEKLEKAKEMLETTQVNLCSQSQSKDSR 2994

RESULT 48
US-09-914-259-66
; Sequence 66, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
; US-09-914-259-66

Query Match 6.4%; Score 97; DB 4; Length 961;
Best Local Similarity 22.3%; Pred. No. 1.6;
Matches 67; Conservative 52; Mismatches 125; Indels 56; Gaps 14;

QY 10 VEVVKSIAETADGALDLYNKLQVLPKTFDETTELKSRPKQEVSOEASVLVDIKVLL 69
Db 666 LEELKQKIST-----LKQNEQLQTAV-----TQVVSQIQKHQKQYNL-----LKVQL 708

QY 70 -MDSQDK--YFEATQTVVWCGVVTQLLSAYILLDFEYNEKAS-----AKDILIR 118
Db 709 GKDSQHQGPYTDGAQ---MNGVQPEISR---LREBIEELKSNRELLOLSQAEKSLIE 761
```

Qy 119 ILDDGVKK--LNEAQKSLTSSQFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKE 176  
Db 762 NLKSSQLSPGTNBOSSATGDSQIAELKOELATLKSQ--NSQSVETIKLQTEKQELLQK 820  
Qy 177 AYAGAAAGIVAGPGLIISYSIAAGVIEGKL-----IPELNNRLKTVQNFFTSLSATVK 230  
Db 821 TEAFKAPVPGSETVIAKTTD--VEGRLSALLQETKELKEIKALSERTAIKEQLD 878  
Qy 231 OANKIDAAKLKLAIEAAGEIKETETTRFFVYDDMLSLKGAAKMINTCNEYQQ 290  
Db 879 SSNSTI-----AILQNEKNKLEVDITDSKEQDQLLV--LLADQDQKIFSLKKNLKE 928

## RESULT 49

US-09-489-039A-8352  
; Sequence 8352, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 8352  
; LENGTH: 350  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-8352

Query Match 6.4%; Score 96.5; DB 4; Length 350;  
Best Local Similarity 22.1%; Pred. No. 0.39;  
Matches 68; Conservative 46; Mismatches 93; Indels 101; Gaps 17;

Qy 39 TFDTEIKELSRFOEYSQASVL---VGD-IKVLLMDSQDKYFEATQTVYEMCGVVTQLL 94  
Db 42 TRDVVMAKVTQLGYRPNANAQALATQVSDTIGVVMDVSDAFGALVKAVD---TVAQQH 98  
Qy 95 SAYILLFDEYNE-KKASQAKDILIR-----ILDD-----GVKKLNEAQ 131  
Db 99 QKYVLIGNSYHEAKEKHAIEVLIRQCSALIVHSKALSDDLSDFMHPGWLINRIV 158  
Qy 132 KSLTSSQFNNASGKLLALDSQLTNDSEKSSYFQSQ---QVDRIRKEAYAGA--AAGI 185  
Db 159 PGVAHRCVGLDNYSGALMA--TRMLLNHGHRIGYLSNHHGIEDDDMRGWSKALQEQGI 217  
Qy 186 VA-----GP-----FGL--IISY--SIAAGVI-----EGKLIP-- 210  
Db 218 IAPDSWIGSGSPDMQGGGEAAWVLLGNLGLTAVFAYNDSMAAGALTTLKDNGIVVPQHL 277  
Qy 211 -----LNNRLKTVQNFFTSLSATVKQAKNDIDAANKLATEIAAIGEI--KTE 256  
Db 278 SLIGFDDIPISRYTDPQLTTRVPVMSMA-----KLATELALLGAAGKLD 322  
Qy 257 TETTRFYV 264  
Db 323 REATHCFM 330

## RESULT 50

US-08-924-629C-6  
; Sequence 6, Application US/08924629C  
; Patent No. 6403082  
; GENERAL INFORMATION:  
; APPLICANT: Stiles, Michael E.  
; APPLICANT: Vederas, John C.  
; APPLICANT: van Belkum, Marius J.  
; APPLICANT: Worobo, Randy W.  
; APPLICANT: Worobo, Rodney J.

; APPLICANT: Greer, G. Gordon  
; APPLICANT: McMullen, Lynn M.  
; APPLICANT: Leisner, Jorgen J.  
; APPLICANT: Poon, Alison  
; APPLICANT: Franz, Charles M.A.P.  
; TITLE OF INVENTION: No. 6403082elBacteriocins, Transport and Vector System and Method  
; FILE REFERENCE: 660.0005US  
; CURRENT APPLICATION NUMBER: US/08/924,629C  
; CURRENT FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: US 60/026,257  
; PRIOR FILING DATE: 1996-09-05  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 6  
; LENGTH: 457  
; TYPE: PRT  
; ORGANISM: lcaad;  
US-08-924-629C-6

Query Match 6.4%; Score 96.5; DB 4; Length 457;  
Best Local Similarity 21.6%; Pred. No. 0.59;  
Matches 73; Conservative 63; Mismatches 113; Indels 89; Gaps 18;

Qy 3 SIFAEOTVEVWKAIB-----TADGALD-----LYNKYLDQVIEPKTFDETIKE 46  
Db 39 SLPAKREI-VKASGHIIPAKVLSDIQSTSNNAIDSNQLTENKVVKGGDTLVFTSGNEK 97  
Qy 47 LSRFKOEYSQASVLVGDIKVL-----LMSQDKYFEATQTVYEMCGVVTQLLSAYILL 100  
Db 98 IS--SOLLTQQLNNDRLKSLDTYKQSI VNGRSEFGTDFQGYD-----SLFNGYMAQ 149  
Qy 101 FD-----EYNEKK-----ASAQKDLIRILDGVKKLNEA---QKSLTSSQS----- 140  
Db 150 VDTLTSEFNQSSDKQTADQQAHHQIDVLKQGSKNQQLANYQAILTSINSNTKPTNPP 209  
Qy 141 ----FNNASGLIALDSQLTNDSEKS-----SYFQSQVDRI--KEAYAGAAAGIV-AGP 189  
Db 210 YQAIYDNYSAQLKS--AQITDDKQVKQTALSNVQQIDQLTSSSYDSQIAGITKSGP 267  
Qy 190 FGLIISYSIAAGVIEGKLI---PELNNRLKTVQNFFTSLSATVKQANKD----- 235  
Db 268 LSQSSTLDKIADLKQQQLASAQKEINDQ-----QQLDELKAKQSSANEDYQDVTVIKAPED 323  
Qy 236 -----IDAAKLKLAIEAIAIGIKTE--TETTRFYVDY 266  
Db 324 GILHLATDKTKIKYFPKGTITIAQIYPKLTOKTALNVEY 361

Search completed: January 5, 2005, 10:59:00  
Job time : 20.0526 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:52:17 ; Search time 17.5576 Seconds  
(without alignments)  
1671.423 Million cell updates/sec

Title: US-09-993-292B-2  
Perfect score: 1515  
Sequence: 1 MTSIFAQTVVVKSAIETA.....NEYQQRHGKKTLEFVDPVAS 305

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Databases : PIR 79.\*  
1: Pirl.\*  
2: Pirl.\*  
3: Pirl.\*  
4: Pirl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description         |
|------------|-------|-------------|--------|----------|---------------------|
| 1          | 1503  | 99.2        | 305    | 2 AE0673 | haemolysin HlyE [i  |
| 2          | 1402  | 92.5        | 305    | 2 C64864 | hemolysin E - Esch  |
| 3          | 1387  | 91.6        | 305    | 2 E90838 | hemolysin E [impor  |
| 4          | 1387  | 91.6        | 305    | 2 E85696 | probable pore form  |
| 5          | 118   | 7.8         | 1496   | 2 T05834 | hypothetical prote  |
| 6          | 116   | 7.7         | 1023   | 1 LEECA  | hemolysin A - Esch  |
| 7          | 115.5 | 7.6         | 622    | 2 T22716 | hypothetical prote  |
| 8          | 115   | 7.6         | 587    | 2 A90394 | bP52 protein homol  |
| 9          | 114.5 | 7.6         | 478    | 2 T12818 | hypothetical prote  |
| 10         | 114.5 | 7.6         | 652    | 2 B59102 | hypothetical prote  |
| 11         | 113.5 | 7.5         | 727    | 2 AC1814 | hypothetical prote  |
| 12         | 113.5 | 7.5         | 1127   | 2 T28317 | ORF MSV156 hypothe  |
| 13         | 113.5 | 7.5         | 2823   | 2 F87908 | protein T22A3.8 [i  |
| 14         | 113.5 | 7.5         | 2823   | 2 T23064 | hypothetical prote  |
| 15         | 113.5 | 7.5         | 3102   | 2 T43291 | laminin alpha chai  |
| 16         | 113   | 7.5         | 956    | 2 S30834 | hypothetical prote  |
| 17         | 113   | 7.5         | 1005   | 2 A64465 | hypothetical prote  |
| 18         | 112.5 | 7.4         | 1024   | 2 S10056 | hemolysin A - Esch  |
| 19         | 112   | 7.4         | 1039   | 2 S18199 | myosin heavy chain  |
| 20         | 111.5 | 7.4         | 1999   | 1 S21801 | myosin heavy chain  |
| 21         | 111.5 | 7.4         | 2819   | 2 A90551 | conserved hypotchet |
| 22         | 111   | 7.3         | 1295   | 2 T24587 | hypothetical prote  |
| 23         | 110.5 | 7.3         | 584    | 2 S75986 | hypothetical prote  |
| 24         | 110.5 | 7.3         | 927    | 2 AG1739 | transmembrane prot  |
| 25         | 110   | 7.3         | 821    | 2 S67087 | hypothetical prote  |
| 26         | 110   | 7.3         | 1938   | 1 MWK11  | myosin heavy chain  |
| 27         | 109.5 | 7.2         | 1492   | 2 T14652 | protein J - Versin  |
| 28         | 109.5 | 7.2         | 1545   | 2 T14966 | phage lambda-relat  |
| 29         | 109   | 7.2         | 595    | 2 F75008 | hypothetical prote  |

|    |       |     |      |          |                    |
|----|-------|-----|------|----------|--------------------|
| 30 | 109   | 7.2 | 739  | 2 H75001 | methyl-accepting c |
| 31 | 109   | 7.2 | 955  | 1 A32554 | leukotoxin A - Pas |
| 32 | 109   | 7.2 | 1098 | 2 B70232 | hypothetical prote |
| 33 | 108.5 | 7.2 | 1066 | 1 A48669 | kinesin-related pr |
| 34 | 108   | 7.1 | 726  | 2 T44825 | hypothetical prote |
| 35 | 108   | 7.1 | 998  | 2 T00227 | hemolysin A toxin  |
| 36 | 108   | 7.1 | 2139 | 2 T18296 | myosin heavy chain |
| 37 | 107.5 | 7.1 | 1292 | 2 D84727 | probable RAD50 DNA |
| 38 | 107.5 | 7.1 | 1875 | 2 S38173 | myosin-like protei |
| 39 | 106.5 | 7.0 | 539  | 2 F72288 | methyl-accepting c |
| 40 | 106.5 | 7.0 | 927  | 2 AH1369 | transmembrane prot |
| 41 | 106.5 | 7.0 | 1272 | 2 C90593 | hypothetical prote |
| 42 | 106.5 | 7.0 | 1473 | 2 A35186 | salivary agglutini |
| 43 | 106.5 | 7.0 | 2346 | 2 T13829 | Tpr homolog - frui |
| 44 | 106   | 7.0 | 598  | 2 I41078 | hemolysin - Escher |
| 45 | 105.5 | 7.0 | 540  | 2 T44967 | gas-vesicle protei |
| 46 | 105   | 6.9 | 1601 | 2 AB1730 | hypothetical prote |
| 47 | 104.5 | 6.9 | 566  | 2 S54091 | hypothetical prote |
| 48 | 104.5 | 6.9 | 1329 | 2 T48959 | kinesin-like prote |
| 49 | 104   | 6.9 | 329  | 1 VMUT21 | VSG expression sit |
| 50 | 104   | 6.9 | 633  | 2 T41332 | casp homolog - fis |

ALIGNMENTS

RESULT 1

AE0673

haemolysin HlyE [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain C)  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AE0673  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AE0673  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-305 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD01758.1; PID:g16502606; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY1498  
C:Superfamily: Escherichia coli hemolysin E

|    |     |   |     |
|----|-----|---|-----|
| Qy | 1   | MTSIFAQTVVVKSAIETADGALDLYNKYLDQVLPWKTFDETIKLSRFKQEQYSQASV       | 60  |
| Db | 3   | MTGIFAEQTVVVKSAIETADGALDLYNKYLDQVLPWKTFDETIKLSRFKQEQYSQASV      | 62  |
| Qy | 61  | LVGDIKVLMLDSQDKYFEATQTVYEWCGVVVTQLLSAYILLDFEYNEKKASQAQKIDILIRIL | 120 |
| Db | 63  | LVGDIKVLMLDSQDKYFEATQTVYEWCGVVVTQLLSAYILLDFEYNEKKASQAQKIDILIRIL | 122 |
| Qy | 121 | DGVKKLNEAQKSLTSSQSFNNASGKLALDLSQTLNDFSEKSYFOSQVDRKKEAYAG        | 180 |
| Db | 123 | DGVKKLNEAQKSLTSSQSFNNASGKLALDLSQTLNDFSEKSYFOSQVDRKKEAYAG        | 182 |
| Qy | 181 | AAAGIVAGPPGLIISYSIAAGVIEGKLIPELNRLKTVQNFFTSLSATVQANKDIDAAK      | 240 |
| Db | 183 | AAAGIVAGPPGLIISYSIAAGVIEGKLIPELNRLKTVQNFFTSLSATVQANKDIDAAK      | 242 |
| Qy | 241 | LKLAETIAAIGEIKTETETTRFYVDYDDMLSLKGAAGKQWINTCNEYQQRHGKKTLEFV     | 300 |
| Db | 243 | LKLAETIAAIGEIKTETETTRFYVDYDDMLSLKGAAGKQWINTCNEYQQRHGKKTLEFV     | 302 |

A:Accession: E90838  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-305 <HAY>  
A:Cross-references: UNIPROT:Q9REB3; GB:BA000007; PIDN:BA815100.1; PID:g13361141; GSPDB:G1  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: EC91677  
C:Superfamily: Escherichia coli hemolysin E

Query Match 91.6%; Score 1387; DB 2; Length 305;  
Best Local Similarity 90.1%; Pred. No. 1.3e-82;  
Matches 273; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTSIFAQTVVWVKSATETADGALDLYNKYLDQVVPWKTFTETIKELSRFKQYSQASV 60  
DB 3 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVVPWKTFTETIKELSRFKQYSQASV 62

QY 61 LVGDIKVLMDSDQKYFEATQTVVWCGVVTQLLSAYILLDFEYNEKASAKDILIRIL 120  
DB 63 LVGNIKTLLMDSQKYFEATQTVVWCGVATQLLAAYILLDFEYNEKASAKDILIKVL 122

QY 121 DGVVKLNEAOKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKRAYAG 180  
DB 123 DGGITKLNEAOKSLLVSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKRAYAG 182

QY 181 AAAGIVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAAK 240  
DB 183 AAAGVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAAK 242

QY 241 LKLTETIAAIGETETETTRFYVDYDDMLSLKGAACKMINTCNEYQORHGKKTLPFV 300  
DB 243 LKLTETIAAIGETETETTRFYVDYDDMLSLKGAACKMINTCNEYQORHGKKTLPFV 302

QY 301 PDV 303  
DB 303 PEV 305

RESULT 4  
E85696  
probable pore forming hemolysin hlyE [imported] - Escherichia coli (strain O157:H7, subsp. C; Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: E85696  
R:Perna, N.T.; Plunkett III, C.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, I.; Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: E85696  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-305 <STO>  
A:Cross-references: UNIPROT:Q9REB3; GB:AE005174; PIDN:AG56033.1; GSPDB:G1  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: hlyE  
C:Superfamily: Escherichia coli hemolysin E

Query Match 91.6%; Score 1387; DB 2; Length 305;  
Best Local Similarity 90.1%; Pred. No. 1.3e-82;  
Matches 273; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTSIFAQTVVWVKSATETADGALDLYNKYLDQVVPWKTFTETIKELSRFKQYSQASV 60  
DB 3 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVVPWKTFTETIKELSRFKQYSQASV 62

QY 61 LVGDIKVLMDSDQKYFEATQTVVWCGVVTQLLSAYILLDFEYNEKASAKDILIRIL 120  
DB 63 LVGNIKTLLMDSQKYFEATQTVVWCGVATQLLAAYILLDFEYNEKASAKDILIKVL 122

QY 121 DGVVKLNEAOKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKRAYAG 180  
DB 123 DGGITKLNEAOKSLLVSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKRAYAG 182

QY 181 AAAGIVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAAK 240  
DB 183 AAAGVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAAK 242

QY 241 LKLTETIAAIGETETETTRFYVDYDDMLSLKGAACKMINTCNEYQORHGKKTLPFV 300  
DB 243 LKLTETIAAIGETETETTRFYVDYDDMLSLKGAACKMINTCNEYQORHGKKTLPFV 302

QY 301 PDV 303  
DB 303 PEV 305

QY 301 PDV 303  
DB 303 PDV 305

RESULT 2  
C64864  
hemolysin E - Escherichia coli (strain K-12)  
N:Alternate names: hemolysin-inducing protein  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: C64864  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: C64864  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-305 <BLAT>  
A:Cross-references: GB:AB000216; GB:U00096; NID:g1787417; PIDN:AAC74266.1; PID:g1787430;  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: hlyE, hpr  
C:Function:  
A:Description: hemolytic activity  
A:Note: pore formation  
C:Superfamily: Escherichia coli hemolysin E  
C:Keywords: cytotoxic; cytotoxin; hemolysis; transmembrane protein  
F:181-197/Domain: transmembrane #status predicted <TMM>  
F:123/Active site: Asp #status predicted

Query Match 92.5%; Score 1402; DB 2; Length 305;  
Best Local Similarity 91.1%; Pred. No. 1.4e-83;  
Matches 276; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

QY 1 MTSIFAQTVVWVKSATETADGALDLYNKYLDQVVPWKTFTETIKELSRFKQYSQASV 60  
DB 3 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVVPWKTFTETIKELSRFKQYSQASV 62

QY 61 LVGDIKVLMDSDQKYFEATQTVVWCGVVTQLLSAYILLDFEYNEKASAKDILIRIL 120  
DB 63 LVGDIKVLMDSDQKYFEATQTVVWCGVATQLLAAYILLDFEYNEKASAKDILIKVL 122

QY 121 DGVVKLNEAOKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKRAYAG 180  
DB 123 DGGITKLNEAOKSLLVSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRIKRAYAG 182

QY 181 AAAGIVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAAK 240  
DB 183 AAAGVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAAK 242

QY 241 LKLTETIAAIGETETETTRFYVDYDDMLSLKGAACKMINTCNEYQORHGKKTLPFV 300  
DB 243 LKLTETIAAIGETETETTRFYVDYDDMLSLKGAACKMINTCNEYQORHGKKTLPFV 302

QY 301 PDV 303  
DB 303 PEV 305

RESULT 3  
E90838  
hemolysin E [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C:Accession: E90838  
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference strain O157:H7  
A:Reference number: A99629; MUID:21156231; PMID:11258796



Db 123 DDGITKLEAQLSVSSQFNASGKLALDSQTNDFSEKSYFOSQVDKIRKAYAG 182  
Qy 181 AAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQNFYTSLSATVQANKDIDAAK 240  
Db 183 AAGVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQNFYTSLSATVQANKDIDAAK 242  
Qy 241 LKLAETAAIGETITETTFYVDYDMLSLGKAQKMINTCNEYQORHGKKTLPYV 300  
Db 243 LKUTTEAAIGETITETTFYVDYDMLSLGKAQKMINTCNEYQORHGKKTLPYV 302  
Qy 301 PDV 303  
Db 303 PEV 305

RESULT 5  
T05634  
hypothetical protein F20D10.190 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C;Accession: T05634  
R;Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.H.  
submitted to the Protein Sequence Database, February 1999  
A;Reference number: Z15420  
A;Accession: T05634  
A;Molecule type: DNA  
A;Residues: 1-1496 <BEV>  
A;Cross-references: UNIPROT:Q95ZK7; EMBL:AL035538  
A;Experimental source: cultivar Columbia; BAC clone F20D10  
C;Genetics:  
A;Map position: 4  
A;Introns: 1042/2; 1212/2; 1232/1; 1263/3; 1349/3  
A;Note: F20D10.190

Query Match 7.8%; Score 118; DB 2; Length 1496;  
Best Local Similarity 19.6%; Pred. No. 11;  
Matches 64; Conservative 67; Mismatches 122; Indels 74; Gaps 13;

Qy 4 IPAQVTVVKSATETADGALDLYNK--YLDQVVPKLT-----FDETIKLSRF---KQ 52  
Db 153 IEAKTVKGMKEMGRDVRDGVVVEEKSQVEELKWKKEQFKHLEAYEYKLNLFKDSKK 212  
Qy 53 EYSQEAIVLGDIKVL--LMSQDKYFEATQTVYEWGCVVTTQLLSAYILLFDEYNEKKA 109  
Db 213 EWEKSKLDIETYSLOTKLDVTRISEDLOKQKQMCNGALTO-----EETRKHIL 263  
Qy 110 SAQKDILIRLDDGVKLENAQKSLTSSQFNASGKLALDSQTNDFSEKSYFQ-- 167  
Db 264 EIQVSEFKAYEDAFACQDARTQ-----DLAKRRDWEVLAELQTLMDQAYFKEM 316  
Qy 168 ----SQVDRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEG---KLIPELNNRLKTVQN 220  
Db 317 KYENGKLEQENRELLGSLK-----ELQETIQQSGNSALSQKKNFRLEN 362  
Qy 221 FFTLSAT-----VKQANKDIDAAKLAETAAIGETITETTFYVDYDMLSLGKAQKMINTCNEYQORHGKKTLPYV 264  
Db 363 IHKNCSANLRSKEAENWSQVEKVEEINDYKLOQSKAAALKEVELENCRSSTAKORL 422  
Qy 265 DYDDLMLSLGKAQKMINTCNEYQOR 291  
Db 423 QYEEISIMFL-----VLSRTVSEAQSR 444

RESULT 6  
LECA  
hemolysin A - Escherichia coli  
C;Species: Escherichia coli  
C;Date: 30-Sep-1988 #sequence\_revision 30-Sep-1988 #text\_change 09-Jul-2004  
C;Accession: A24433; I41280  
R;Felmler, T.; Pellett, S.; Welch, R.A.  
J. Bacteriol. 163, 94-105, 1985  
A;Title: Nucleotide sequence of an Escherichia coli chromosomal hemolysin.

A;Reference number: A24433; MUID:85234404; PMID:3891743  
A;Accession: A24433  
A;Molecule type: DNA  
A;Residues: 1-1023 <FEL>  
A;Cross-references: UNIPROT:P09983; GB:M10133; GB:M12863; NID:G146377; PIDN:AAA23975.1;  
A;Experimental source: strain J96, O4 serotype  
R;Stanley, P.; Packman, L.C.; Koronakis, V.; Hughes, C.  
Science 266, 1992-1996, 1994  
A;Title: Fatty acylation of two internal lysine residues required for the toxic activity  
A;Reference number: A5387; MUID:95099325; PMID:7801126  
A;Contents: annotation; lysine palmitoylation  
A;Note: lysine modification is performed by the hlyC gene product  
R;Haertlein, M.; Schiesl, S.; Wagner, W.; Rdest, U.; Kreft, J.; Goebel, W.  
J. Cell Biol. 22, 87-97, 1983  
A;Title: Transport of hemolysin by Escherichia coli.  
A;Reference number: I41280  
A;Accession: I41280  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1,'T',3,'V',5,'T',7-44 <RES>  
A;Cross-references: GB:M29173; NID:G146337; PIDN:AAA23957.1; PID:G146338  
C;Genetics:  
A;Gene: hlyA  
C;Function:  
A;Description: attacks blood cell membranes and causes cell lysis  
A;Superfamily: hemolysin A; hemolysin A homology  
C;Keywords: calcium binding; cytolysis; exotoxin; hemolysis; lipoprotein; tandem repeat;  
F;246-791/Domain: hemolysin A homology <HLYA>  
F;723-851/Region: 9-residue repeats (G-G-X-G-[DN]-D-X-[LVITYF]-X)  
F;563,689/Binding site: palmitate (Lys) (covalent) #status experimental

Query Match 7.7%; Score 116; DB 1; Length 1023;  
Best Local Similarity 20.5%; Pred. No. 9.2;  
Matches 69; Conservative 59; Mismatches 111; Indels 98; Gaps 11;

Qy 2 TSIFAEQVWVK-----SAIETADGALDLYNKYLDQVVPKLTDETIKLSRFKQE 53  
Db 155 TALSSWKIDELIKKQSGNVSSSELAKASIELINQLVDTA-----ASLNNVNSFSQ 207  
Qy 54 YSQEASVL-----VGDIKVLMSQDKYFEATQTVYEWGCVVTTQLLSAYILL-FDEY 104  
Db 208 LNKLGSLVSLNTHLGVGN-KLQNLPLNDINAGLDTV---SGILSAISASFILSNADAD 263  
Qy 105 NEKKAQAKDILIRLDD---GVKLENAQKSLTSSQFNASGKLALDSQTNDFSE 161  
Db 264 TGTGAAGVELTTKVLGNVGKISQVYIAQRAAQLSTS----- 302  
Qy 162 KSSYFQSQVDRIRKEAYAGAAAGIVAGPGLIIS---YSIAAGVIEGKLIPELNNRLKT 217  
Db 303 -----AAAGLIASVTVLAIISPLSFLSIADKFKRANKIEYSQRFKK 344  
Qy 218 VQNFPTLSATVQANKDIDAAKLAETAAIGETITETTFYVDYDMLSLGKA 277  
Db 345 LGYDGSLLAAPHKETCAIDASLTRISLVASVSSGISAATTSV---GAPVSALVGA 400  
Qy 278 -----AKWINTCNEYQORHGK 294  
Db 401 VTGIISGLEASKQAMFEHVASKVADVIAEWKKGK 437

RESULT 7  
T22716  
hypothetical protein F55C5.8 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T22716  
R;Harris, B.  
submitted to the EMBL Data Library, August 1996  
A;Reference number: Z19604  
A;Accession: T22716  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-622 <WIL>

A;Cross-references: UNIPROT:Q20822; EMBL:Z78198; PIDN:CAB01573.1; GSPDB:GN00023; CESP:FP  
A;Experimental source: clone F55C5  
C;Genetics:  
A;Gene: CESP:F55C5.8  
A;Map position: 5  
A;Introns: 24/3; 44/2; 147/3; 251/3; 561/3

Query Match 7.6%; Score 115.5; DB 2; Length 622;  
Best Local Similarity 21.2%; Pred. No. 5.2;  
Matches 68; Conservative 56; Mismatches 109; Indels 87; Gaps 15;  
Qy 41 DETIKELSRFKQBYQASVNLGDIKLLMDSQDKFEATQTYVCGVVTQLSAY--- 97  
Db 251 DKLISEM-RASATSAEVVTTWGGAKSTVDDE-----KAKQVVQVQVTEVELAQOCTPK 304  
Qy 98 --ILLFDEYNEKKASAKDILIRLDDGVKLANE-AOKSLTSSQSF-----NNASGKL 148  
Db 305 EKMALF-----EKATADTRDAIDRISDIIRKSSENADTTVLQSIKAYLEFLKMGTAARY 360  
Qy 149 LALDSQLTNDFSEKSYFQS-----QVDRIRKEAYAG 180  
Db 361 LAI---IDNTKSEKSKPQDILLRLDYSVIEYKEVABIPGADHDKNLIQAFVKVEYRA 417  
Qy 181 AAGAGVAGPGLIISYSTAAGVI-----EGKLIPELNNRLKTVQNFFTSLSATVK 230  
Db 418 FRCFYMASSYALHKYSEAAALFDRTVSRVQDAEGKL-----KKLKSSSFITNETQSSLN 472  
Qy 231 QANKDIDAAK--LKLATEIAAIGIKETETRTTFYVDVDDMLSLGAKAKMINTCNEY 288  
Db 473 ELRSEVESAKVTVRAARLASAAGDVKTDSELAKI-ID-----KRPLETVTNEW 519  
Qy 289 QQ---RHGKTLFEVDDVAS 305  
Db 520 RQWDRVNSLKKKTTIP-VAS 538

RESULT 8  
A90394  
bPS2 protein homolog (bps2) [imported] - Sulfolobus solfataricus  
C;Species: Sulfolobus solfataricus  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C;Accession: A90394  
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayer, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A;Description: Sulfolobus solfataricus complete genome.  
A;Reference number: A99139  
A;Accession: A90394  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-587 <KUR>  
A;Cross-references: UNIPROT:Q97WH8; GB:AE006641; NID:g13815540; PIDN:AAK42408.1; GSPDB:G  
C;Genetics:  
A;Gene: bps2

Query Match 7.6%; Score 115; DB 2; Length 587;  
Best Local Similarity 20.3%; Pred. No. 5.2;  
Matches 61; Conservative 66; Mismatches 103; Indels 70; Gaps 14;  
Qy 10 VEVVKSIAETADGALDLYNKYLDQVTPWKTFFDETI--KELSRFKQBYQASVNLGDIKV 67  
Db 42 ISLTSSIK---AEDLLNVFADSGYVVEALDNKLYRRIKRNGLGSEKNLIMDDRA 97  
Qy 68 LLMSQDKYFRATQTVVCGVVTQLSA-----YTLFDEYNEKKASAKDILIRLDD 121  
Db 98 LLL-----TYFSPENR-----LVTQLSGDGNVEWFTTTSKINEIK--AKKELOKLT 145  
Qy 122 DGVKKLNEAQSLLTSSQSFNN---ASGKLALDSQLTNDFSEKSYFQSOVDRIKRAY 178  
Db 146 AEINARDELQK-----KYNIREIOAKIRAID-----BEIDKLEKE-- 181  
Qy 179 AGAAAGVAGPGLIISYIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVKQANKDIDA 238

Db 182 RESSNIVAK-----TTYTIT--LTRQNKINELINKVKQDELANLLEFALKKIEEIQN 234  
Qy 239 AKLKLATEIAAIGIKETETRTTFYVDYDD-----LMLSLLGAAKGM-----INTCN 286  
Db 235 KESKVSPIKTOLEKEMEEINEKLUKLTNDRSELELKVLRVLEEVNESDRHLDTCN 294  
RESULT 9  
Ti2818  
hypothetical protein yonD - Bacillus subtilis phage SPBc2  
C;Species: Bacillus subtilis phage SPBc2  
C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
C;Accession: Ti2818; F69913  
R;Lazarovic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.  
submitted to the EMBL Data Library, August 1997  
A;Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 prophage  
A;Reference number: Z17583  
A;Accession: Ti2818  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-478 <LAZ>  
A;Cross-references: UNIPROT:O64067; EMBL:AF020713; NID:g3025478; PID:g3025532; PIDN:AAACI  
K;Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleri  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A;Authors: Lauber, J.; Lazarovic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, Y.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtilis.  
A;Title: The complete genome sequence of the Gram-positive bacterium  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: F69913  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-478 <KUN>  
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14031.1; PID:el183560,  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: yonD

Query Match 7.6%; Score 114.5; DB 2; Length 478;  
Best Local Similarity 22.6%; Pred. No. 4.3;  
Matches 57; Conservative 41; Mismatches 73; Indels 81; Gaps 12;  
Qy 25 DLYNKY-LDOVIPWKTFFDETIKELSRFKQBYQASVNLGDIKLLMDSQDKYFEATQTV 83  
Db 268 DVTDTYFVNVYWS--DENSVD-KYFKNYIR-----TGDTVSIDFSKTEVF-----MT 315  
Qy 84 YEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRLDDGVKLANEAKSLTSSQSFNN 143  
Db 316 RNWEEVPEPIQSOL-----NQDEQIK---DLTKQVNLNKKDKVIGEQFNT 360  
Qy 144 ASGKLALDSQ-----LTNDFSEKSYFQSOVDRIKRAYAGAAAGIVAG 188  
Db 361 ASEKLVQLNSEVQOLKEKYEKHEKTLLEQKLEKSEKFEYKAKFEALNAEE----- 409  
Qy 189 PFGLIISYIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVKQANKDIDAOKLATEIA 248  
Db 410 -----KFS-----TEEVQNL---IHASVKQ-DEGEKAVIQNTMLV 442  
Qy 249 AIGELKTEITET 260  
Db 443 DLVSVPTEITNT 454

**RESULT 10**

B59102  
hypothetical protein pX01-90 - Bacillus anthracis virulence plasmid pX01  
C;Species: Bacillus anthracis  
C;Date: 12-Nov-1999 #sequence\_revision 12-Nov-1999 #text\_change 09-Jul-2004  
C;Accession: B59102  
R;Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler, J. Bacteriol. 181, 6509-6515, 1999  
A;Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harboring genes encoding the protective antigen, edema factor, and lethal toxin.  
A;Reference number: A59091; MUID:99445483; PMID:10515943  
A;Accession: B59102  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-652 <UKI>  
A;Cross-references: UNIPROT:Q9X360; GB:AF065404; NID:g4894216; PIDN:AAD32394.1; PID:g4894216  
A;Experimental source: strain Sterne  
A;Note: similar to hypothetical protein; PF0765w (980 aa); Plasmodium falciparum (AE001159)  
C;Genetics:  
A;Gene: pX01-90  
A;Genome: plasmid

Query Match 7.6%; Score 114.5; DB 2; Length 652;  
Best Local Similarity 18.6%; Pred.No. 6.4;  
Matches 58; Conservative 66; Mismatches 106; Indels 81; Gaps 10;

QY 25 DLYNKYLDOVIPWKTFTDETIKELSRFKQYSQBASVL-----VGDIKVLMLDSQDYKE 78  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 237 DVINOKIDE---FDKLNSQRKDLERMLELNQLKSOLKKQSPOQLDKNNKLKESQRSLLLE 293

QY 79 ATQ-----TVVEWCQVVTQLLSAYILLDFEYNE--KKASAQKDILIRILDGVKK 126  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 294 LNFKDSNRLEINSEIKLNDRKAEALLSLIMELIKQQSEFDDKLIKNEKDLINKREDLINR 353

QY 127 LNEAQK-----SLLTSOSFNNAASKLALDSOLTDFSEKSSYFQ---SQV 170  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 354 IASSELAKKABNLTKLVLFKVQBALNKSQYLYYINKLDNELRELADTKYNSDNKI 413

QY 171 DRIRKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVK 230  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 414 SRLKNH-----IGEYNKQLEKIEN-----ELE 435

QY 231 QANKDIDAACKLAT-EIAAIGEIKETETETRYFYVDYDDL-----MLSLKGAAKQKINT 284  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 436 ECKNKIDNTKKLAEPDKSNKKQOESELSVLQNLKIDELGKRRHRQELEASQKKALDE 495

QY 285 CNEYQQRHGKK 295  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 496 AKKINKKLAEK 506

**RESULT 11**

AC1814  
hypothetical protein all10059 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C;Accession: AC1814  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, H.; Tanaka, K.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.; Nakazaki, N.; Shimojo, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AC1814  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-727 <KUR>  
A;Cross-references: UNIPROT:Q8ZON7; GB:BAO000019; PIDN:BAB77583.1; PID:g17135037; GSPDB:G17135037  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all10059

Query Match 7.5%; Score 113.5; DB 2; Length 727;  
Best Local Similarity 22.2%; Pred.No. 8.6;

|         |     |              |     |            |      |        |     |      |     |
|---------|-----|--------------|-----|------------|------|--------|-----|------|-----|
| Matches | 75; | Conservative | 58; | Mismatches | 134; | Indels | 71; | Gaps | 15; |
|---------|-----|--------------|-----|------------|------|--------|-----|------|-----|

  

|    |     |   |     |
|----|-----|---|-----|
| Qy | 5   | FABQTVEVVKSAIETADGALDLNKKYLDQVIPWKTFDETIKLSRFRKQEQYSQBSAVLVGD | 64  |
| Db | 194 | FINARLPVIKQQLRQAERKLEQPRKKHNLDPQLSQVLTK-KTQEQQLTRTQLQD        | 252 |
| Qy | 65  | IKVLLMDSQDKYEATQTVYEWGVVYQLL---SAYLLFDEYNEKKKASAKDILIRILD     | 121 |
| Db | 253 | INSRYQNLEIRITEASQK-----ALVSMRLAQSSRYKLTLSLQKTEQSLAKEQL-RYTD   | 306 |
| Qy | 122 | DG--VRKLNKSAKSLIT-SQSFNNAAG-----KLALDLSQLTNDF-----            | 159 |
| Db | 307 | DSPIVQSLKQRRSRLTVRQELKRLTAELKTTQTEPEQLVGVDPNLVEEFVQVQTAL      | 366 |
| Qy | 160 | ----SEKSSYFQSDVRIIRKEAVAGAAAGVAGPGFLIISYSIAAGVIEGKLIPELNRL    | 215 |
| Db | 367 | GLIANEKS--LRSEQRIRSE-----LSKYPSLIARYQ-----RLAUFVETQR          | 407 |
| Qy | 216 | KTVQNFF-TLSLTVKQAKNDIDAAKLKLAETAAIGETETETTRTFYVDYDDL-----     | 269 |
| Db | 408 | KTLEQMLAQOQSGMGKIAHGFPD---WLVEEANLG---TYVGNDRFMLVFGVMITGPI    | 460 |
| Qy | 270 | ---WLSLLGAAKMTNCNEYQORHGKKTLFVDPVA                            | 304 |
| Db | 461 | LGVLALIMGMRHRIIHSAQDLQKVNRLRLGTVPKLA                          | 498 |

  

|   |
|---|
| RESULT 12   |
| T28317  |
| ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus         |
| C/Species: Melanoplus sanguinipes entomopoxvirus                                |
| C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004     |
| C/Accession: T28317   |
| J. Virol. 73, 533-552, 1999   |
| R. Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.        |
| A>Title: The genome of Melanoplus sanguinipes entomopoxvirus.                   |
| A/Reference number: Z20484; MUID:99102612; PMID:9847359                         |
| A/Accession: T28317   |
| A/Status: preliminary; translated from GE/EMBL/DBJ                              |
| A/Molecule type: DNA  |
| A/Residues: 1-1127 <AFO>  |
| A/Cross-references: UNIPROT:Q9YV76; EMBL:AF063866; NID:G4049647; PIDN:AAC97677. |
| C/Genetics:   |
| A/Note: MSV156  |

|                       |        |                  |                 |                     |
|-----------------------|--------|------------------|-----------------|---------------------|
| Query Match           | 7.5%;  | Score 113.5;     | DB 2;           | Length 1127;        |
| Best Local Similarity | 19.2%; | Pred. No. 15;    |                 |                     |
| Matches               | 60;    | Conservative 56; | Mismatches 106; | Indels 91; Gaps 12; |

  

|    |     |   |     |
|----|-----|---|-----|
| Qy | 25  | DLNKKYLDQVIPKMTFDETIKELSRKQFVSQBSVLVGDIKV-LLMDSODKFEATQTV     | 83  |
| Db | 214 | DELNKLDE--SKKEFIKQEEELNKTKTIDRKQEEELIKKLNDEKFNFIDEKQKLLD---QI | 268 |
| Qy | 84  | YEWCGVVTQLLSAYILLFDEYNEKKASAKDIL-----IRLLDGVKKLNEAOKSLTSS     | 138 |
| Db | 269 | NSKINTLNENIKGMVLYETKNNKISNLQNEILLKNDSTIKSLDEKQKLLDELQKNNIT    | 328 |
| Qy | 139 | QSFNNASGKLL-----ALDSOLT-----NDFSEKSSYFSQSVDRIRKEAYAGAAAGI     | 185 |
| Db | 329 | SLANKSNTKITNIQQLLESSLTDFFNNANININELKSKIKLFDNDIOKLNND-----     | 379 |
| Qy | 186 | VAGPFGLLISYSTAAGVIEGKLIPELNNRLKTVQNEFTSLSATVKQANKDIDAAUKLKAT  | 245 |
| Db | 380 | -----ITEQNNK-----ITDFNNSTRIPKE-----KLDI                       | 404 |
| Qy | 246 | EIAAIGEIKTET-----ETTRFY-----VDYDDL-----LSLLKGAAKMI            | 282 |
| Db | 405 | EYKKIDDIKNNNLQKLEESYKKIDEQTEYYKNNKEYNDIIELKNNNLQKLEENKKID     | 464 |
| Qy | 283 | NTCNEYQQRHGKK   | 295 |
| Db | 465 | EOTEYYKKNKINS   | 477 |

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19980  
A;Accession: T25096  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2823 <W12>  
A;Cross-references: EMBL:Z81125; PIDN:CA03385.1; GSPDB:GN00019; CESP:T22A3.8  
A;Experimental source: clone T22A3  
C;Genetics:  
A;Gene: CESP:T22A3.8  
A;Map position: 1  
A;Introns: 45/1; 282/2; 312/3; 416/2; 1255/3; 1329/3; 1418/3; 1776/2; 1988/2; 2760/2  
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.5%; Score 113.5; DB 2; Length 2823;  
Best Local Similarity 20.8%; Pred. No. 49;  
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;

QY 25 DLYNKYLDOVIPW-KTFDETIKELSRFKQESVQESVLVG-----DIKVLMD 72  
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADAERMSLVGKRINRYKEVSEIEKLAVEA 1939  
QY 73 QDK-----YFEATQTV-----YEWCGVVTQLLSAYILLFDEY 104  
DB 1940 EDQIAYSRSIEKARSEELMNFEDKEKINMTLAELPDLVEQCNITLL---YSQLIDEY 1996  
QY 105 NEK-----KASAKDILIRILD-----DGVKLNLEAQS 133  
DB 1997 DEEYVQTAGRHAKEVQAQK-IVDRFVDTRTETENPLKASHAYENIVEALKNATEAVDS 2055  
QY 134 LITSSQSFNAGKLLALDSQLTNDFSEKSSYFQSDVRIRKEAYAGAAAGIVAGPGLI 193  
DB 2056 AAEASE-----AVSKMLGSGESGDANES--LRSOLEKLNESLSN-----2097  
QY 194 ISYSIAAGVIE--GKLIPELNNRLKTQVNFPTSLSATVQKANDIDAKKLATEIAAIG 251  
DB 2098 VDNSNAVKIVEELKEKDLTDLRGLHNLKTSI-----VKRLG 2136  
QY 252 EIKTETETFRFYVDYDMLSLKGAKKM-INTCNEYQORHGKKT 296  
DB 2137 VIKNEASS---WDDKHDRMHSILKNGAKTAHRSANVKESGIKT 2179

RESULT 15

T43291  
laminin alpha chain - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T43291  
R;Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang,  
submitted to the EMBL Data Library, June 1998  
A;Description: Expression, function and evolution of laminin alpha chains.  
A;Reference number: 222397  
A;Accession: T43291  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-3102 <ZHU>  
A;Cross-references: UNIPROT:O45614; EMBL:AF074902; PIDN:AAC26793.1  
C;Genetics:  
A;Map position: 1  
A;Note: lamal/2  
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.5%; Score 113.5; DB 2; Length 3102;  
Best Local Similarity 20.8%; Pred. No. 56;  
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;

QY 25 DLYNKYLDOVIPW-KTFDETIKELSRFKQESVQESVLVG-----DIKVLMD 72  
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADAERMSLVGKRINRYKEVSEIEKLAVEA 1939  
QY 73 QDK-----YFEATQTV-----YEWCGVVTQLLSAYILLFDEY 104

RESULT 13

F87908  
protein T22A3.8 [imported] - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C;Accession: F87908; E87908  
R;anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A;Reference number: A75000; MUID:99069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A;Accession: F87908  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2823 <STO>  
A;Cross-references: UNIPROT:O45614; GB:chr\_1; PIDN:CAAL5432.1; PID:G3924779; GSPDB:GN000  
A;Accession: E87908  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2823 <ST2>  
A;Cross-references: GB:chr\_1; PIDN:CA03385.1; PID:G3924881; GSPDB:GN00019; CESP:T22A3.8  
C;Genetics:  
A;Gene: T22A3.8  
A;Map position: 1  
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.5%; Score 113.5; DB 2; Length 2823;  
Best Local Similarity 20.8%; Pred. No. 49;  
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;

QY 25 DLYNKYLDOVIPW-KTFDETIKELSRFKQESVQESVLVG-----DIKVLMD 72  
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADAERMSLVGKRINRYKEVSEIEKLAVEA 1939  
QY 73 QDK-----YFEATQTV-----YEWCGVVTQLLSAYILLFDEY 104  
DB 1940 EDQIAYSRSIEKARSEELMNFEDKEKINMTLAELPDLVEQCNITLL---YSQLIDEY 1996  
QY 105 NEK-----KASAKDILIRILD-----DGVKLNLEAQS 133  
DB 1997 DEEYVQTAGRHAKEVQAQK-IVDRFVDTRTETENPLKASHAYENIVEALKNATEAVDS 2055  
QY 134 LITSSQSFNAGKLLALDSQLTNDFSEKSSYFQSDVRIRKEAYAGAAAGIVAGPGLI 193  
DB 2056 AAEASE-----AVSKMLGSGESGDANES--LRSOLEKLNESLSN-----2097  
QY 194 ISYSIAAGVIE--GKLIPELNNRLKTQVNFPTSLSATVQKANDIDAKKLATEIAAIG 251  
DB 2098 VDNSNAVKIVEELKEKDLTDLRGLHNLKTSI-----VKRLG 2136  
QY 252 EIKTETETFRFYVDYDMLSLKGAKKM-INTCNEYQORHGKKT 296  
DB 2137 VIKNEASS---WDDKHDRMHSILKNGAKTAHRSANVKESGIKT 2179

RESULT 14

T23064  
hypothetical protein T22A3.8 - Caenorhabditis elegans (fragment)  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T23064; E87908  
R;Barlow, K.  
submitted to the EMBL Data Library, October 1997  
A;Reference number: Z19669  
A;Accession: T23064  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2823 <W1L>  
A;Cross-references: UNIPROT:O45614; EMBL:AL008595; PIDN:CAAL5432.1; GSPDB:GN00019; CESP:  
A;Experimental source: clone H10E24  
R;McMurray, A.

Db 1940 EDQIAYSRNIEKARSEELNMNFKEDKINWTLAELPDLVEQCQNTLL---YSQLIDEY 1996  
QY 105 NEK-----KASAKQDILIRILD-----DGVKKLNEAQS 133  
Db 1997 DEBYVOTAGHABKLEVQAOK-IVDRPVRTTETENPLKASHAYENIVEALKNATEAVDS 2055  
QY 134 LRTSSQSFNNASGKLALDSQNTDFSEKSSYFQSQVDRIKAYAGAAAGIVAGPGLI 193  
Db 2056 AAEASE-----AVSKMLGSESGDANEEES--URSQLEKUNESSLN-----2097  
QY 194 ISYISIAAGVIE--GKLIPELNNRLKTVQNFFTSLSATVVKQAKDIDAALKLATEIAAIG 251  
Db 2098 VDNNAVKIIVEELKKEKKOLTDRLGHLNELKTSI-----VKRLG 2136  
QY 252 EIKTETETTRFYVDYDMLSLKGAQKM-INTCNEYQORHGKKT 296  
Db 2137 VIKNEASS---WDDKDRMHSILKNGAKTAHRSANYKSESEGIKT 2179

RESULT 16  
S30834  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-May-1993 #sequence\_revision 28-May-1993 #text\_change 09-Jul-2004  
C:Accession: S30834; S50501  
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor, submitted to the EMBL Data Library, February 1993  
A:Reference number: S30834  
A:Accession: S30834  
A:Molecule type: DNA  
A:Residues: 1-956 <MUL>  
A:Cross-references: UNIPROT:P32618; GB:U18779; EMBL:L10830; NID:G603625; PIDN:AAB64999.1  
R:Dietrich, F.S.  
A:Description: The sequence of *S. cerevisiae* cosmids 8199, 8334, and 9871.  
A:Reference number: S50491  
A:Accession: S50501  
A:Molecule type: DNA  
A:Residues: 1-956 <DIE>  
A:Cross-references: EMBL:U18779; NID:G603625; PIDN:AAB64999.1; PID:G603636; MIPS:YEL043W  
C:Genetics:  
A:Cross-references: SGD:S0000769  
A:Map position: 5L

Query Match 7.5%; Score 113; DB 2; Length 956;  
Best Local Similarity 24.4%; Pred. No. 13;  
Matches 68; Conservative 45; Mismatches 78; Indels 88; Gaps 15;

QY 38 KTFDETIKLSRPFQKQYSQASVLVGDIVKVLMD-----SQDKYFEATQTVYEWCGVITQ 92  
Db 272 KSLKSNIKSL-----ENSKLLTDLKTEKLNKIDKSKKIKRNMNMQKWSQEDTE 322  
QY 93 LLS-----AYILDFBYNEKSAOKDI-----LIRILDGVKKLNEAOKSLTS---S 138  
Db 323 LLSKDTIKYFKLLNESNASVANINKEIESLQNEISKMEESKRLNASKSLITIVVN 382  
QY 139 QSFNN-----ASGKLALDSQNTDIFS-EKSSYF-----QSQVDRIKAYEA 179  
Db 383 ANVENDKPIASGELSAVLKGL-NDFTLEKNGFLSNAGEEFLSKLNADSSLIKMIKQE---438  
QY 180 GAAAGIVAGPGLIISYIAA-----GVIEGKLIP-ELNNR-LKT-----V 218  
Db 439 -----LSIDQELANWKLQSRNLLKKISALENFNEMSLNNRLNLTKLAVQPY 486  
QY 219 QNFPTLSAT-----VKQAKNDIDAALKLATEIAAIGEI 253  
Db 487 KNGKDSLAATNSNNAEKNRSSGSIQLPLSNMNSRTGSI 525

RESULT 17  
A64465  
C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C:Accession: A64465  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, I.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinatock, K.G.; Merrick, J.M.; Glodek, A.; rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Frazer, C.M.; Smith, H.O.; Woese, C. A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii* A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: A64465  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1005 <BUL>  
A:Cross-references: UNIPROT:Q58718; GB:U67572; GB:L77117; NID:G1591958; PIDN:AAB99331.1  
C:Genetics:  
A:Map position: REV1273394-1270377  
C:Superfamily: hypothetical protein MJ1322

Query Match 7.5%; Score 113; DB 2; Length 1005;  
Best Local Similarity 19.7%; Pred. No. 14;  
Matches 60; Conservative 55; Mismatches 91; Indels 98; Gaps 12;

QY 7 EQTVVVKSAIETADGALDLYNKYLQVLPWKTFDE---TIKELSRF-----KQYSQ 56  
Db 605 DELEIDIKSQLNFKP---NFYQYLSAVSYLSNVDEGIRNRRIKEIENIVSGWNKEKCRE 661  
QY 57 EASVL-----VGDIKVLMDSDQKFEATQTVYEWCGVITQSLLSAYILLDFEYNEKKAS 110  
Db 662 ELNKLREDEREINRLKDLNKLNLKKE-----LIEIENRRSLKFDKYKEYLGL 710  
QY 111 AQDILIRILDGVKKLNEAOKSLTSSQSFNNASGKLALD-----152  
Db 711 TEKLEELKNIKDLEEI-----YINCNSKILALDNIKRYNKEDIYLYNNK 757  
QY 153 ----SOLTNDFFSEKSSYFQSQVDRIKAYAGAAAGIVAGPGLIISYIAAGVIEGKLI 208  
Db 758 ILEVKNKEINDIERISYINQKLE-----INYNEE-----EHKKI 792  
QY 209 PEL--NNR--LKTQVNFHTSLSATVVKQAKNDIDAALKLATEIAAIGETETTRFYV 264  
Db 793 KELYENKQELDNVREKQTEIETGIEYLVKQDVESLKARL-----KEMSNLEKEKEKLTKFV 848  
QY 265 DYDD 268  
Db 849 EYLD 852

RESULT 18  
S10056  
hemolysin A - *Escherichia coli* plasmid phly152  
C:Species: *Escherichia coli*  
C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 04-Mar-2000  
C:Accession: S10056  
R:Hess, J.; Wells, W.; Vogel, M.; Goebel, W. FEMS Microbiol. Lett. 34, 1-11, 1986  
A:Title: Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparison A:Reference number: S07209  
A:Accession: S10056  
A:Molecule type: DNA  
A:Residues: 1-1024 <HES>  
A:Cross-references: EMBL:M14107  
C:Genetics:  
A:Genome: plasmid phly152  
C:Superfamily: hemolysin A; hemolysin A homology  
C:Keywords: lipoprotein  
F;247-792/Domain: hemolysin A homology <LYA>  
F;564,690/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.4%; Score 112.5; DB 2; Length 1024;  
Best Local Similarity 20.8%; Pred. No. 15;  
Matches 70; Conservative 55; Mismatches 115; Indels 97; Gaps 11;

QY 2 TSIFAEQTVVWV-----SAIFADGALDLYNKYLQVDPWPKTFDETIKELSRFQOE 53  
Db 155 TALSSMKIDELIKKQSGGNVSSSELAKASIELINQLVDVTASLN-----NNVNSFSQO 208  
QY 54 YSQEASVL-----VGDIKVLLMDSQDKYFEATQVYEWCVVQVLLSAYILL-FDEY 104  
Db 209 LNTLGSVLSNKHNGVN-KJQNLPLNDNIGAGLTV---SGILSAISAFILSNADAD 264  
QY 105 NEKASAKQDILIRIIDD---GVKKLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSE 161  
Db 265 TRTKAAGVELTTKVLGNCKGISQYIIAQRAGLST----- 303  
QY 162 KSSYFQSQVDRIRKEAYAGAGVAGVPFGLIIS---YSIAAGVIEGKLIPELNNRLKT 217  
Db 304 -----AAAAGLIASAVTLAISPLSIFLSIADKFRANKIEYSQRFK 345  
QY 218 VQNFPTLSATVKQANKDIDAAKILATEIAAIGEIKTETETTRFYVDYDDLMLSLKGA 277  
Db 346 LGYDGSLLAFPHKETGALDASLTITSTVLASVSSGISAAATTSV---GAPVSALVGA 401  
QY 278 -----AKQMINTCNEVQORHGK 294  
Db 402 VTGIISGILEASKQAMFEHVASKMADVIAEWEKHGK 438

RESULT 19  
S18199  
myosin heavy chain - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C:Date: 22-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 02-Feb-2001  
C:Accession: S18199  
R:Stewart, A.F.R.; Canoret-Mercado, B.; Perlman, D.; Gupta, M.; Jakovcic, S.; Zak, R.  
J. Mol. Evol. 33, 357-366, 1991  
A:Title: Structural and phylogenetic analysis of the chicken ventricular myosin heavy chain  
A:Reference number: S18199; MUID:92130260; PMID:1774788  
A:Accession: S18199  
A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1039 <SITE>  
A:CROSS-references: EMBL:X59552; NID:G62995; PID:CAA42130.1; PID:G62996  
A>Note: In the authors' translation 45-Lys is shown after residue 40, and, consequently,  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: ATP; nucleotide binding; P-loop  
F:402-409/Region: nucleotide-binding motif A (P-loop)

Query Match 7.4%; Score 112; DB 2; Length 1039;  
Best Local Similarity 19.9%; Pred. No. 17;  
Matches 69; Conservative 67; Mismatches 143; Indels 68; Gaps 11;

QY 1 MTSIFAEQTVVVKSAIETADGALDLYNK--YLDQVDPWKT-FDETIKELSRFQOEY 57  
Db 369 MTFMNDLTTQTKLQSENGEFVRQLEEKESLISQLSRGKTSFTQOIEELRRQLEETKS 428  
QY 58 ASVLVGDIKVLLMD-----SQDKYFEATQVYEWCVVQVLLSAYILLPPEYNEK 107  
Db 429 KNALAHQAARHDCDLLEQVEEEOFAKAEQRLAKGNVAQWRTYETDAIQRTEE 488  
QY 108 KASAKQDILIRI-----LDDGVKLINEAQKSLTSSQSFNNASGKLLAL 151  
Db 489 LEDAKKLLARLQAEAEATEAANKCSLEKAKHRLQNEQEDMDIDLEKANSAAA---SL 545  
QY 152 D-----SQNTNPFSEKSVFQSQVDRIRKEAYAGAGVAGVPFGLIISYSIAAGVIEG 205  
Db 546 DKQGRGFDKIINDWKQYEESSQAELEASQKE-----ARSLSTELFLKKNAYETDLHLR- 599  
QY 206 KLIPELNNRLKTQVNFPTLSATVKQANKDI-----DAAKILATEIAAIGE 253  
Db 600 -----TLKRNKNLQOEESIDLTNQISEGNKNLHIEKVKQVEQEKSEVQLALEEABGALE 655  
QY 254 KTETETTRFYVDYDDLMLSLKGAQKQMINTCNEYQO---RHGKTL 297  
Db 656 HEESKTLRF-----QLELSQLKADFERKLAEDKQEMQNIIRNQORTI 697

RESULT 20  
S21801  
myosin heavy chain, neuronal [similarity] - rat  
N:Alternate names: myosin II  
N:Contains: myosin ATPase (EC 3.6.4.1)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C:Accession: S21801; PNO013; S18134  
R:Sun, W.; Chantler, P.D.  
J. Mol. Biol. 224, 1185-1193, 1992  
A:Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain  
A:Reference number: S21801; MUID:92235856; PMID:1569576  
A:Accession: S21801  
A:Molecule type: mRNA  
A:Residues: 1-1999 <SUN>  
A:CROSS-references: UNIPROT:Q63731; EMBL:X62659  
R:Sun, W.; Chantler, P.D.  
Biochem. Biophys. Res. Commun. 175, 244-249, 1991  
A:Title: A unique cellular myosin II exhibiting differential expression in the cerebral  
A:Reference number: PNO013; MUID:91151356; PMID:1998509  
A:Accession: PNO013  
A:Molecule type: mRNA  
A:Residues: 1914-1998, 'I' <SU2>  
A:Experimental source: brain  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide  
F:174-181/Region: myosin motor domain homology <MMOT>  
F:84-763/Domain: myosin motor domain homology <MMOT>  
F:541-575/Region: nucleotide-binding motif A (P-loop)  
F:653-675/Region: actin binding #status predicted  
F:836-1999/Domain: coiled coil #status predicted <COI>  
F:836-1276/Region: S2  
F:1277-1999/Region: light meromyosin  
F:125/Modified site: NS,NG,N6-trimethyllysine (Lys) #status predicted  
F:180/Binding site: ATP (Lys) #status predicted  
F:693-703/Active site: Cys #status predicted  
F:1916/Binding site: phosphate (Ser) (covalent) #status predicted  
F:1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 7.4%; Score 111.5; DB 1; Length 1999;  
Best Local Similarity 19.6%; Pred. No. 42;  
Matches 60; Conservative 49; Mismatches 114; Indels 83; Gaps 11;

QY 7 EQTVVVKSAIE---TADGALDLYNKYLQVDP--WKTFFDT---IKELSRFQOEY 57  
Db 1165 EQVNLKKTLEAEAKTHEAQIOEMRQKHSQAEVLEAEQETKRKVANLEKAKOTLENE 1224  
QY 58 ASVLVGDIKVLLMDSQDKYFEATQVYEWCVVQVLLSAYILLFDEYNEKASAKQDILI 117  
Db 1225 RGEALANEVYLLQGGRD-----SHKKKKVEAQLOELQ 1257  
QY 118 RLDDGVKLINEAQKSLTSSQSFNNASGKLLALD---SQNTNPFSEKSVFQSQVDRIR 174  
Db 1258 VKFNEGERRVTELDADKTKQLQVELDNTVGLSSQDSKSLTKDFSALESQLDQTQELQ 1317  
QY 175 KEAYAGAGVAGVPFGLIISYSIAAGVIEGKLIPELNNRLKTQV---NFTSLSATVK 230  
Db 1318 EEN-----RQKL--SLSTKLQVEDEKNSFRQLEEE 1349  
QY 231 QANKDIDAAKILATEIAAIGEIKTETETTRFYVDYDDL-MLSLLKGAQKQMINTCNEYQ 289  
Db 1350 EAKINLEK---QIATLHAQVADMKKQME-----DSVCLTAEEVKKQLQKDLGLS 1398  
QY 290 QRHGCK 295  
Db 1399 QRHEEK 1404

RESULT 21  
A90551  
conserved hypothetical protein MYPU\_3130 [imported] - Mycoplasma pulmonis (strain UAB C)  
C:Species: Mycoplasma pulmonis

```

398 LIEAVINAKQULEIKLIAEENSBIEEFKIEVAQRVGIGAVAIADUSQNRISJIVFSDNN 435
      271 LSLIKGAAKTQI 282
      ||| |||
      :| | |
458 LALQGNTAPYML 469

RESULT 24
AG1739

```



|    |                       |  |               |       |             |      |
|----|-----------------------|--|---------------|-------|-------------|------|
|    | Query Match           | 7.3%;  | Score 110;    | DB 2; | Length 821; |      |
|    | Best Local Similarity | 20.4%;   | Pred. No. 17; |       |             |      |
|    | Matches               | 68;  | Conservative  | 52;   | Mismatches  | 109; |
|    |                       |  |               |       | Indels      | 104; |
|    |                       |  |               |       | Gaps        | 12;  |
| QY | 10                    | VVVKSAIETADGALDINYKYLDOVTPWKTFTETIKELSRFKQEVSQEASVLVGDIKVL-  | 68            |       |             |      |
|    |                       | : : :    : : :    : : :    : : :    : : :                    |               |       |             |      |
| Db | 373                   | VEKKPKRIKELNTEIKVLNS--NQKILOEKFPASITEVNHIIKEHENTVTILOQNEKILN | 430           |       |             |      |
|    |                       | : : :    : : :    : : :    : : :    : : :                    |               |       |             |      |
| QY | 69                    | -----LMSODKYFEATQVYEWCGVTTQLLSAYILLFDEYNKKASAQKDIL           | 116           |       |             |      |
|    |                       | : : :    : : :    : : :    : : :    : : :                    |               |       |             |      |
| Db | 431                   | DKNVELENKKAEKGNDKLSEVETINDLNSRIVOL-----NDKIEST--DIV          | 477           |       |             |      |
|    |                       | : : :    : : :    : : :    : : :    : : :                    |               |       |             |      |
| QY | 117                   | IRILDGVKYLNEAQAQSLLTSSQSFNNASKULLALDSQLTNDFSEKSSYFQSOVDRIKE  | 176           |       |             |      |
|    |                       | : : :    : : :    : : :    : : :    : : :                    |               |       |             |      |
| Db | 478                   | LASKENELDNKLISKETLSIKDFNDSD-----LIQINELISTKN--LOQRMD----     | 526           |       |             |      |
|    |                       | : : :    : : :    : : :    : : :    : : :                    |               |       |             |      |
| QY | 177                   | AVAGAAAGIVAGPGLIISVISIAGVIEGKLIPELAN-----BLKTQNPFPTSLSATVKQA | 232           |       |             |      |
|    |                       | : : :    : : :    : : :    : : :    : : :                    |               |       |             |      |
| Db | 527                   | -----DLNNLNDNLKVQDCLKIKNEETLKUK                              | 553           |       |             |      |

Db 554 EAEIDSLNEMDELKQKITSDDFKWQSKYETVEAKIRNAEVTLNGDIEDLKESK 613

Qy 270 -----MLSLLGKAARKMINTCNEYQQRHGKKTL 297  
          ::      |      |      |      |      |  
Db 614 LHLLETTITELKNVHKLNECELEKQKFETSL 646

RESULT 26

MWKM1  
myosin heavy chain D [similarity] - Caenorhabditis elegans  
N:Alternate names: myosin heavy chain I  
N:Contains: myosin ATPase (EC 3.6.4.1)  
C:Species: Caenorhabditis elegans  
C>Date: 28-Feb-1986 #sequence revision 19-May-2000 #text\_change 09-Jul-2004  
C:Accession: T21193; T23973; S02772; A02993  
R:McMurray, A.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19388  
A:Accession: T21193  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1938 <MW>  
A:Cross-references: UNIPROT:P02567; EMBL:Z71261; PIDN:CAA95806.1; GSPDB:GN000011  
A:Experimental source: clone F21C3  
R:Gardner, A.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19825  
A:Accession: T23973  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A;Accession: R06C7  
A;Experimental source: clone R06C7  
R;Dibo, N.J.; Naruyana, I.N.; Krause, M.; Karn, J.  
J. Mol. Biol. 205, 603-613, 1989  
A;Title: Sequence analysis of the complete *Caenorhabditis elegans* myosin heavy  
A;Reference number: S02771; MUID:89178677; PMID:2926820  
A;Accession: S02772  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-376, 'V', 378-390, 'V', 392-576, 'L', 578-680, 'I', 682-1938 <DIB>  
A;Residues: 1-376, 'V', 378-390, 'V', 392-576, 'L', 578-680, 'I', 682-1938 <DIB>  
A;Cross-references: EMBL:X08065; NID:56785; PIDN:CAA30854.1.; PID:56786  
R;Karn, J.; Brenner, S.; Barnett, L.  
Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983  
A;Title: Protein structural domains in the *Caenorhabditis elegans* unc-54 myosin  
A;Reference number: A93958; MUID:83273600; PMID:6576334  
A;Accession: A02993  
A;Molecule type: DNA  
A;Residues: 24-93, 'E', 95-97, 'R', 99-376, 'V', 378-388, 'GDV', 392-407, 'N', 409-473,  
C;Genetics:



Db 924 IAEEREARVEGDKANAKQIEAMKSSVDDSVAAVEEMKKTVAEVERASAEASTNIEALAKT 983

Qy 152 -----DSQLTND-----FSEKSSYFQSQVDRIKKEAYAGAAAGIVAGPF 190

Db 984 NIDLALRQDEQHKQMVNNAKIATTKTFADDMSAMASKVEIRAE----- 1029

Qy 191 GLIISYSIAAGVIEGKLIPELNNRLKTVQNFTSLSATVKQANKOIDAA 239

Db 1030 ---IGEDIRASILE-----ETARVEADKTIATHISKLEAQLNDDDISAA 1070

RESULT 28

Tl14966

phage lambda-related host specificity protein J - Yersinia pestis plasmid pMT1

C:Species: Yersinia pestis

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C:Accession: Tl14966

R:Lindler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.

A:Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIMS plasmid

A:Reference number: Z18268; MUID:99043898; PMID:9826348

A:Accession: Tl14966

A:Status: preliminary; translated from GB/EMBL/DBDJB

A:Molecule type: DNA

A:Residues: 1-1545 <L>

A:Cross-references: UNIPROT:Q9ZH03; EMBL:AF074611; NID:G3883003; PID:G3883049; PIDN:AAC8:

C:Genetics:

A:Gene: Y1049

A:Genome: plasmid pMT1

Query Match 7.2%; Score 109.5; DB 2; Length 1545;

Best Local Similarity 19.7%; Pred. No. 41;

Matches 57; Conservative 54; Mismatches 101; Indels 77; Gaps 9;

Qy 2 TSIAEQTVVEVKSIAETADGALDLYNKYLDQVIPWKTETIKELSRFKQEQYSQASVL 61

Db 849 SSELGQELLEIDA--KASDAQDAVAINKQMBESL--KELDQSVADLDLSKLEDTSGRIEQV 904

Qy 62 VGDIKVLIMDSQDKYFEATQTVYEWGCVVTQL-----SAYIL 99

Db 905 QNDLKNEVSGTLKDKNVDALQVEDSNAALVELQETVSEQKAIAGAVEAAHAALDNASAL 964

Qy 100 LFEYNEK-----KASA-QKDILIRLDGVGKLNBAQKSLTSSQSFNNASGKLAL--- 151

Db 965 IAEEREARVEGDKANAKQIEAMKSSVDDSVAAVEEMKKTVAEVERASAEASTNIEALAKT 1024

Qy 152 -----DSQLTND-----FSEKSSYFQSQVDRIKKEAYAGAAAGIVAGPF 190

Db 1025 NIDLALRQDEQHKQMVNNAKIATTKTFADDMSAMASKVEIRAE----- 1070

Qy 191 GLIISYSIAAGVIEGKLIPELNNRLKTVQNFTSLSATVKQANKOIDAA 239

Db 1071 ---IGEDIRASILE-----ETARVEADKTIATHISKLEAQLNDDDISAA 1111

RESULT 29

F75008

hypothetical protein PAB1294 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004

C:Accession: F75008

R:anonymous, Genoscope

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure

A:Reference number: A75001

A:Accession: F75008

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-595 <KAW>

A:Cross-references: UNIPROT:Q9UYA4; GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB5050

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1294

QY 262 FYVDYDDMLSLKGAACKMINTCNEYQQRHGK 294  
 DB 704 SAIEQQTAAIEELRAAQELKDMVGRMRQIVGK 736

RESULT 31

A35254  
 leukotoxin A - Pasteurella haemolytica (serotype T10)  
 A:Alternate names: lktA protein  
 C:Species: Pasteurella haemolytica  
 C:Date: 10-Aug-1990 #sequence\_revision 15-Nov-1996 #text\_change 09-Jul-2004  
 C:Accession: S37145; A35254, S34237; S34235  
 R:Rains, A.F.; Aitchison, K.; Donachie, W.  
 submitted to the EMBL Data Library, September 1993  
 A:Description: DNA sequence of the leukotoxin A gene from P. haemolytica T10 serotype.  
 A:Reference number: S37145  
 A:Accession: S37145

A:Molecule type: DNA  
 A:Residues: 1-955 <LA1>  
 A:Cross-references: UNIPROT:P51117; EMBL:Z26247; NID:G400424; PIDN:CAA81206.1; PID:G4004.  
 R:Highlander, S.K.; Engler, M.J.; Weinstein, G.M.  
 J. Bacteriol. 172, 2343-2350, 1990  
 A:Title: Secretion and expression of the Pasteurella haemolytica leukotoxin.  
 A:Reference number: A35254; MUID:9023688; PMID:2185213  
 A:Accession: A35254  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 950-955 <HIG>  
 A:Cross-references: GB:M24197; GB:M34943; GB:M34944  
 R:Lainson, A.F.; Aitchison, K.D.; Donachie, W.  
 submitted to the EMBL Data Library, June 1993  
 A:Description: DNA sequence of the carboxy terminal end of leukotoxin A from the T3 seroi  
 A:Reference number: S34235  
 A:Accession: S34237  
 A:Molecule type: DNA  
 A:Residues: 745-955 <LA2>  
 A:Cross-references: EMBL:Z22884; NID:G311828; PIDN:CAA80498.1; PID:G311829  
 A:Experimental source: serotype T3  
 A:Accession: S34235  
 A:Molecule type: DNA  
 A:Residues: 723-955 <LA3>  
 A:Cross-references: EMBL:Z22887; NID:G311824; PIDN:CAA80501.1; PID:G311825  
 A:Experimental source: serotype T10  
 C:Function:  
 A:Description: attacks cell membranes and causes cell lysis  
 C:Superfamily: hemolysin A, hemolysin A homology  
 C:Keywords: calcium binding; cytotoxicity; exotoxin; hemolysis; lipoprotein; tandem repeat;  
 F:240-786/Domain: hemolysin A homology <HL1A>  
 F:718-809/Region: 9-residue repeats (G-G-X-G-[DN]-D-X-[LVIVF]-X)  
 F:718-726/Region: repeat  
 F:727-735/Region: repeat  
 F:736-744/Region: repeat  
 F:745-753/Region: repeat  
 F:754-762/Region: repeat  
 F:763-771/Region: repeat  
 F:772-780/Region: repeat  
 F:781-789/Region: repeat  
 F:792-800/Region: repeat  
 F:801-809/Region: repeat  
 F:556/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.2%; Score 109; DB 1; Length 955;  
 Best Local Similarity 18.3%; Pred. No. 24;  
 Matches 68; Conservative 63; Mismatches 139; Indels 102; Gaps 11;

QY 6 AEOIVVVKSAIETADGALDLYNKVDQVTPKTFDETIKELSRFKQEQSASVLVG-- 63  
 DB 89 AQTSLGTLQNVIGLTERGIVLSAPQLDKLQKNKVGQALGSSSESTAQNFSAOKTVLSGVQ 148  
 QY 64 -----DIKVLMDSDQY-----FEATQTVVWCGVVTQLLSAYILLFDEYNEK 107  
 DB 149 GNSRTVLGMDLDEALQNESDQLTAKAGLELTNSLIENANSVOTLDAFSEIQSQFGSK 208

Query Match 7.2%; Score 109; DB 2; Length 595;  
 Best Local Similarity 23.0%; Pred. No. 13;  
 Matches 71; Conservative 55; Mismatches 109; Indels 74; Gaps 18;

QY 24 LDLYNKVLDQVTPKTFDETIKELSRFKQEQSASVLVG-DIKVLMDSDQKFEATQT 82  
 DB 45 LDLASESLDQIEQSNSTSY-ELSRILNKTREDAVIYKRGITKVDLVPFLKLSG 103  
 QY 83 VYEWCGVVTQLLSAYILLFDEYNEK-----ASAQKDLIRLDDGVKKLNEAQSLLTSS 138  
 DB 104 I-----GNILKRSFMYSNFKNEMPELITAMEESLI-IMDDGIK-----EAKLALIRIS 155  
 QY 139 QSFNNASGKLALD-SQTLNDFSEKSYFOSQVDRIRKEAYAGAAIV-----AGPFGI 192  
 DB 156 KIEFSGAGKLDVNDILEKANKLERESSYHKILERL-KSRVTYPRKGLILFVSNANPYVL 214  
 QY 193 --IISYSTAAG-----VIEKGLPELNRLKTVQNFETSL-----SATVKQANK 234  
 DB 215 ENVTYFGAKNASSVQIILINGSILK-----ANVDNGFFSVNYSFPLPGIYSAVAKSGNL 268  
 QY 235 D-----IDAALKL-----LATBI-AAIGETKTTETTR-----FVYDDDLMLSL 273  
 DB 269 TSNVTIINVSKRKTFFVVPSEVCGRIGETVVISGFLRDLNLGYGVPGRKEIVIDY----- 321

QY 274 LKGAAKQMI 282

DB 322 -KGKATKLI 329

RESULT 30

H75001  
 methyl-accepting chemotaxis protein (tlpc) PAB1336 - Pyrococcus abyssi (strain Orsay)  
 C:Species: Pyrococcus abyssi  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
 C:Accession: H75001  
 R:anonymous, Genoscope  
 submitted to the EMBL Data Library, July 1999  
 A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru  
 A:Reference number: A75001  
 A:Accession: H75001  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-739 <KAW>  
 A:Cross-references: UNIPROT:Q3UYF8; GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CAB5045  
 A:Experimental source: strain Orsay  
 C:Genetics:  
 A:Gene: PAB1336

Query Match 7.2%; Score 109; DB 2; Length 739;  
 Best Local Similarity 19.2%; Pred. No. 17;  
 Matches 64; Conservative 71; Mismatches 138; Indels 60; Gaps 11;

QY 1 MTSIFAQTVVVKSA--TETADGALDLYNKVLDQVTPKTFDETIKELSRFKQEQSSEA 58  
 DB 425 VTETFRSISGLVEMANDLEKRANLAQVSKDVTEAL--NQVNEALQVSIQAQOQETI 482  
 QY 59 SVLVGDIKVLMDSDQKFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAAQDI--- 115  
 DB 483 NEITDGRRLVAQTS-----ESVRAMEBFGAVTEVVS-----IANEGSKGDEALKRIEDI 534  
 QY 116 --LIRILDGKVLKNEAQS-----LTSQSQFNASSGKLLALDS----- 153  
 DB 535 QHMSRIETVSKVAENSRNIEETNVITSIAEQTN-----LLALNAAIBAAARGAGRGF 590  
 QY 154 -----QLTNDPSEKSYFOSQVDRIR---KEAVGAAGI--VAGPFGIISYSIAAG 201  
 DB 591 AVAQETRKLAEBESQKADNKLIIIDKLTIEKAEVATEKGVSVIGESSETLRDITGYL 650  
 QY 202 VIEKGLPELNRLKTVQNFETSLSATVQKANKDIDAAKUKLATEIAAIGETKTTETTR 261  
 DB 651 ANIATLLQETSERITVKE-----QIVRTOEVDKALRLENLAASAEETTASAEVS 703

QY 108 -----KSAQKDILIRILDGCVKLNKAEQAQSKLLTSSQSFNNASG 146  
Db 209 LQNVKGLGALGDKLKTGGDLKAGLGVKSRLLSGATAALVLADKDASTAKR-----VG 263  
QY 147 KLLALDSQLTDFSEK-SSYFQSQVDRIKAEVAGAAAGIVA-GPFLGIISYSIAAGV-- 202  
Db 264 AGFELANQVGNITKAVSSYILAQ-----RVAAGLSSTGPVAALIASTAVAVAI 313  
QY 203 IEKGLPELNRLKTVQNF-----TSLSATVKQANKOIDAAKGLKLAETIAAI-- 250  
Db 314 LSPAGIADKEDRAKSLNAYERFKLGYEGDLSLAIEYQHGTTGDASVTAINALAAIAG 373  
QY 251 -----GEIKTETETTFYVDYDMLSLKGAKKMINTCNE 287  
Db 374 GVSAAAAGSVASPIALLVSGITGVITLQYK-----QAMPEHVANKIHKIVE 424  
QY 288 YQORHGKKTLPF 299  
Db 425 WEXNNGGKNYFE 436  
RESULT 32  
B70232  
hypothetical protein BBG10 - Lyme disease spirochete plasmid G/lp28-2  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 09-Jul-2004  
C:Accession: B70232  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kertlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: B70232  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1098 <KLE>  
A:Cross-references: UNIPROT:O50733; GB:AB000786; NID:G2690008; PIDN:AAC66075.1; PID:G269  
A:Experimental source: strain B31  
C:Genetics:  
A:Genome: plasmid  
Query Match 7.2%; Score 109; DB 2; Length 1098;  
Best Local Similarity 20.1%; Pred. No. 28;  
Matches 81; Conservative 64; Mismatches 127; Indels 130; Gaps 18;  
QY 2 TSIFAEOT-----VEVVKSAITADGALDLYNKY-----LDQV 34  
Db 537 SKIFDQUTDENKKILGVKSVNEFNNSYDFVNEYQNLKESRREIRIITLPHTDQV 596  
QY 35 IPWKTDFETIKELSR-FKQYSQASVLVGDIK--VLLMDSQDKYFEATQ----- 81  
Db 597 SALQKLNDEINENKNAFVKYKGSFETLINESNQVVALEQVNEYKTAIDRSFVEAQK 656  
QY 82 -----TVYWCWGVVTVLLSAYILLFDEYNEKKAQKILIRI-----LDD---GVK 125  
Db 657 ALQKEITDLEW---ETMLLPA-----KERASAEKQWASKIQAMKFKVDEHKSQFK 704  
QY 126 KLNKAEQK-----SLTSSQSFNNA-----SGKLLALDS-----QL 155  
Db 705 KLNKNETNNTIKQVAEKAQDTKSLYDSMIDGLNFKNAFMKIDAGKFLNKDGTGESIGEEF 764  
QY 156 TNPFSKSSYFQSQVDRIKAEVAG-----AAAGIVAGPFLGIISYSIAA-----GVI 203  
Db 765 HNLINGKDVNWGELEKWTQMYESWTKGLKTAAGAVFGPWGEAVELINGLTDLFWVGIL 824  
QY 204 EGKLIPELNRLKTVQNF-----TSLSATVKQANKOIDAAKGLKLAETIAAI-----GEIKTETE 258  
Db 825 KG-----QEKARIKAE-----KKRDELEEKRSSEVELKLEDRFDEIKMRKE 870  
QY 259 TTFYVDYDMLSLKGAKKMINTCNEYQOR-HGKKTLPF 299

Db 871 KLSELDDEYTKETIEFLKQAKSQKQISGESEFQKRLHDVQTEYK 912  
RESULT 33  
A48669  
kinesin-related protein KLP61F - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: A48669; B41298  
R:Heck, M.M.S.; Pergira, A.; Pesavento, P.; Yannoni, Y.; Spradling, A.C.; Goldstein, L.S.  
J. Cell Biol. 123, 665-679, 1993  
A:Title: The kinesin-like protein KLP61F is essential for mitosis in Drosophila.  
A:Reference number: A48669; MUID:94043448; PMID:8227131  
A:Accession: A48669  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1066 <HEC>  
A:Cross-references: UNIPROT:P46863; GB:U01842; NID:G416040; PIDN:AAA03718.1; PID:G416041  
R:Stewart, R.J.; Pesavento, P.A.; Moser, D.N.; Goldstein, L.S.B.  
Proc. Natl. Acad. Sci. U.S.A. 88, 8470-8474, 1991  
A:Title: Identification and partial characterization of six members of the kinesin superfamily: ATP; cell division, mitosis; nucleotide binding; P-loop  
F:20-362/Domain: kinesin motor domain homology <KMT>  
F:103-110/Region: nucleotide-binding motif A (P-loop)  
Query Match 7.2%; Score 108.5; DB 1; Length 1066;  
Best Local Similarity 21.6%; Pred. No. 30;  
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;  
QY 24 LDLYNKYLQDVI PWKTFDETIKELSRFKQYSQASVLVGDIKVLLMDSQDKYFEATQTV 83  
Db 159 LELYNEELCDLL---STDDTTK--IRIFDSTKKGSVIIQGLEIPVHSKDDVYKLEKG 213  
QY 84 YEWCVVTVLLSAYILLFDEYNEKKAQKILIRILDGCVK-----KLNKAEQKSLTSS 138  
Db 214 KERKTTATTLNNA-----QSSRSHTVFSIVVHIRENGIEGEDMLKIKGLNLVDLAGS 265  
QY 139 QSFNNAAGK-----LLALDSOLTN--DFSEKSSYFQSQVDRIKAEVAGAA 182  
Db 266 ENVSKAGNEKGI RVTNINQSLTLGRVITALVDRAHPVYRESKLTLLQESLGRT 325  
QY 183 -----AGIVAGPFL-----IISYSIAAGVIEGKLIPELNRL--KTVQNFSTLSATVK 230  
Db 326 KTSIIATISPCHKDIBETLSTLEYAHRANKIQNK--PEVNQKLTKTTLVKEYTE---EID 380  
QY 231 QANKIDDAKGLKLAETIA--AIGEIKTETETTFYVDYDMLSLK 275  
Db 381 KLKRDLMARDKNGIYLAETTYGEITLKLESQNLRELKMLLLKALK 427  
RESULT 34  
T44825  
hypothetical protein wzc [imported] - Acinetobacter lwoffii  
N:Alternate names: protein tyrosine kinase  
C:Species: Acinetobacter lwoffii  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T44825  
R:Nakat, D.; Gutnick, D.L.  
submitted to the EMBL Data Library, July 1999  
A:Description: Genomic organization of the wzc region of Acinetobacter lwoffii  
A:Reference number: Z24856  
A:Accession: T44825  
A:Status: preliminary; translated from GB/EMBL/DBDJB  
A:Molecule type: DNA



QY 272 SLLKGAKKMINTCNEYQQRHGK 294  
Db 1284 EKLKAKADLEAQLNEQVDNHEK 1306

RESULT 37  
D84727  
Probable RAD50 DNA repair protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: D84727  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shen, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: D84727  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1292 <STO>  
A;Cross-references: UNIPROT:Q9SL02; GB:AE002093; NID:94263721; PIDN:AAD15407.1; GSPDB:GN  
C;Genetics:  
A;Gene: At2g31970  
A;Map position: 2  
C;Superfamily: RAD50 protein

Query Match 7.1%; Score 107.5; DB 2; Length 1292;  
Best Local Similarity 17.2%; Pred. No. 44;  
Matches 69; Conservative 77; Mismatches 138; Indels 117; Gaps 13;

QY 2 TSFAQTVEVVKSAITAGALDLYNKYLQVWPWT-FDE----- 42  
Db 281 STLFKEQQRQYAALPEENED-----TTEELKEWKSFEERLALGTRKRMEREMVD 332  
QY 43 ---TIKELSRFKQEYSOEAASVL--VGDIKVLMDSDQKYFEATQTVYEWGVVTTQLLSAY 97  
Db 333 TETTISLHNAKTYNMLEISKQTEAEHMLLNKRDSTTQNTFFHYNLGNVSTFPSTE 392  
QY 98 ILLFDEYNKKAQAQKDLIRILDGQVKLNEAQSLLTSQSFNNASGKLLALDSQ--- 154  
Db 393 VVL--NLTNRIKSRIGLEMDLLDK--KKSNET--ALSTAWDCYMDANDRWKSIEAQKRA 446  
QY 155 -----LTDNDFEKSYSY-----FQSQ 169  
Db 447 KDEIKMGISKRIEKEITERDSFEFEISTVDVKQTDEREKQVQVELERKTKONSGRPFESK 506  
QY 170 VDRIRKEAYA-----GAAAGIVAGPFGLLISYSI-----AAGVIEGKLIPELNNR- 214  
Db 507 IEQKHIEIYSLEHKIKITLNRERDVMGDAEDRLLTRIDECKDRIRGVILKGLPPEKDMKR 566  
QY 215 -----LKTQVQFTSLSATVQKANKOITDAAKKLATEIAAIGETETTRTTFRYVDYD-- 267  
Db 567 EIVQALRSIEREYDDLSKSRAEKEVNMQLQIEVNNSLFKHKNKDTESRKRYIESKLQ 626  
QY 268 -----DLMSLLKGAKKMINTCNEYQQRHGKTLPE 299  
Db 627 ALKQESVTIDAYPKLLESKAKDRDRKREYNMANGRMQFEE 667

RESULT 38  
S38173  
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)  
N;Alternate names: protein YKR095w; protein YKR415  
C;Species: Saccharomyces cerevisiae  
C;Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 09-Jul-2004  
C;Accession: S38173; S40647; S31207  
R;Baladron, V.; Bailesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo,  
submitted to the Protein Sequence Database, March 1994  
A;Reference number: S38158  
A;Accession: S38173  
A;Molecule type: DNA  
A;Residues: 1-1875 <BAL>

A;Cross-references: UNIPROT:Q02455; EMBL:Z28320; NID:9486586; PID:9486587; MIPS:YKR095w  
A;Experimental source: strain S288C  
R;Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Jin  
Yeast 9, 1349-1354, 1993  
A;Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromosome  
A;Reference number: S40644; MUID:94205265; PMID:8154186  
A;Accession: S40647  
A;Molecule type: DNA  
A;Residues: 1-1875 <BOU>  
A;Cross-references: EMBL:X73541; NID:9450550; PIDN:CAA51948.1; PID:9450554  
A;Experimental source: strain S288C  
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.  
Mol. Gen. Genet. 237, 359-369, 1993  
A;Title: A new yeast gene with a myosin-like heptad repeat structure.  
A;Reference number: S31207; MUID:93247549; PMID:8483450  
A;Accession: S31207  
A;Molecule type: DNA  
A;Residues: 1-300, 'A', 302-1875 <KOE>  
A;Cross-references: EMBL:L01992; NID:9171958; PIDN:AAA34783.1; PID:9171959  
C;Genetics:  
A;Gene: SGD:MLP1  
A;Cross-references: SGD:S0001803; MIPS:YKR095w  
A;Map position: 11R

Query Match 7.1%; Score 107.5; DB 2; Length 1875;  
Best Local Similarity 19.9%; Pred. No. 71;  
Matches 64; Conservative 49; Mismatches 127; Indels 81; Gaps 12;

QY 7 EQTVVVKVSAIETADGALDLYNKYLQVWPWT-FDETI-----KELSRFKQEYSOEAAS 59  
Db 966 EDKISLLKQMFNLNDELLOKGMK--EKADFKRISILQNNKEVEAVKSEYSEKLS 1023  
QY 60 VLVDGIKVLMDSDQKYFEATQTVYEWGVVTTQLLSAYILLFDEYNKKAQAQKDI---- 115  
Db 1024 KIQND-----LDQTTIYANTQNNYE-----QELQKHADVSKTISELR 1061  
QY 116 -LIRILDGQVKLN---BAQKSLTSSQSFNNASGKLLA---LDSQLTWDFSEKSSVFQ 167  
Db 1062 EQLHTYKQGVKTVNLNRDOLNALKENKSWSSQKSLLEQLDLSNRIEDLSQNKLLY 1121  
QY 168 SQVDRIRKEAYAGAAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNRKTQVQFTSLSA 227  
Db 1122 DQI-----QIYTAADKEVNNSTNG-----PGLNLLITLRERDILT 1159  
QY 228 TVQKANKIDAAKKLATEIAAIGETETTR-----FYVDYDDL-----LSLL 274  
Db 1160 KVTVAERDAKMLRQKISLMDVELQDARTKLDNSRVEKENHSSIIQQHDDIMEKLNQLNL 1219  
QY 275 KGAKKMINTCNEYQQRHGK 295  
Db 1220 R---ESNITLRNLENNNNKK 1237

RESULT 39  
F72288  
methyl-accepting chemotaxis protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C;Accession: F72288  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: F72288  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-539 <ARN>  
A;Cross-references: UNIPROT:Q9X0N0; GB:AE001772; GB:AE000512; NID:94981693; PIDN:AAD3622  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM1146

C;Superfamily: methyl-accepting chemotaxis protein

Query Match 7.0%; Score 106.5; DB 2; Length 539;  
Best Local Similarity 19.4%; Pred. No. 17;  
Matches 59; Conservative 63; Mismatches 147; Indels 35; Gaps 10;

QY 10 VEVVKSIAETADGAL---DLNKKYLDQVLPWKTFDETIKELSRFKQKYSQEAASVLVGDIK 66  
DB 206 IEKIRSKDEIGKAAVAKEIKLEILLDIITINKASSEVSSSELSATSEELSANVNSIS 265  
QY 67 VLLMDSQDKFEATOTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKILIRLD----- 121  
DB 266 EALVSLNKEADENSATLEETASIEELSSY-----ADSNKSAQAAMLESTQRVHEQVKS 320  
QY 122 -DGVKKLNEAQKSLTSSQFNNASGKLALDSQLT-----NDFSEKSY--FQSQVDR 172  
DB 321 TERIREITERKAHSTREMSSENTKQALNRLLSMAENINSIVDTINSIAEOTNLLALNAIEA 380  
QY 173 IRKAYAGAAAGIVAGPGLIISYIAA---GVIEGKLIPELNNRLKTQVNEFTSL--S 226  
DB 381 AR-AGEAGRGPAVVADEIRKLAESKARTQOIGELIKLDEINNSKIVESTASALEET 439  
QY 227 ATVQANKDIDAACKLATEIAAIGETETETTRFYVDYDDMLSLKGAAKMINTCN 286  
DB 440 ASLVESIKDV-FESIRIAME-----DVQSRVESVAASTQQSASLELSAGVTRLTELLN 493  
QY 287 EYQQ 290  
DB 494 KTRF 497

RESULT 40

AH1369  
transmembrane protein [imported] - Listeria monocytogenes (strain EGD-e)  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C;Accession: AH1369  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Faihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AH1369  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-927 <GLA>  
A;Cross-references: UNIPROT:O8Y4S2; GB:NC\_003210; PIDN:CAD00438.1; PID:gl6411848; GSPDB:  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo2360

Query Match 7.0%; Score 106.5; DB 2; Length 927;  
Best Local Similarity 20.5%; Pred. No. 33;  
Matches 78; Conservative 44; Mismatches 131; Indels 127; Gaps 16;

QY 1 MTSIFAEQTVVVKSAIETADG---ALDLYNKYLDQVLPW---KTFDETIKELSRFKQE 53  
DB 160 VTKSYAEAFDKTK---ESGDGPAQAADSGKIKGLVKSGQGNKTIISNLKTLADSSLT 216  
QY 54 YSQEASVLI-----VGDIK-----VLLMDS 72  
DB 217 FKDGANTLEVGLKTYTDGVNTAAAGGDKLNAGVSTLAAAGVPLKDGVAALDGGATKLASG 276  
QY 73 QDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKILIRLDGKVKLNEAQK 132  
DB 277 VSTYTSQVDTL---AGGINQAYTGSTALSGLNKNMGS-----VFTLASGITQLNNGQK 327  
QY 133 SLTSSQSFNNASGKLIA-----LDSQLTN-----DFSEKSYFQSQVDPRIRKAYAGAA 182  
DB 328 SLATGLSLVDGSKLSAGLKELDGNLTDKQKIAQLKQGMNDLQQGIDQLNQSVNGEDA 387

QY 183 AGIVAGPGLIISYIAAGVIEGKLIPELNNRL-----KTVQNEFTSLSAT 228  
DB 388 A-----LAKQLAT---LQKSLSDLQGLTFIKSNANFADAEAIKSKINATAGVSAE 434  
QY 229 VKQ-----ANKDIDAACKLATEIAAIGETK-----ETETTRFYVDYDDMLSL 273  
DB 435 DKQKIIDAQADLDEKTKSATQATVATVEQLQSGLSGLDLAAIQTQVT-----EL 483  
QY 274 LKGAAKMINTCNBYQQRHG 293  
DB 484 QTGVAKISAG-----YQAVHG 499

RESULT 41

C90593  
hypothetical protein MYPU 6510 [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
C;Species: Mycoplasma pulmonis  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C;Accession: C90593  
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galieson, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001  
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis.  
A;Reference number: A99512; MUID:21267165; PMID:11353084  
A;Accession: C90593  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1272 <KUR>  
A;Cross-references: UNIPROT:Q98PR8; GB:AL445566; PID:g14090066; PIDN:CAC13824.1; GSPDB:GN  
A;Experimental source: strain UAB CTIP  
C;Genetics:  
A;Gene: MYPU\_6510  
A;Genetic code: SG3

Query Match 7.0%; Score 106.5; DB 2; Length 1272;  
Best Local Similarity 23.1%; Pred. No. 50;  
Matches 52; Conservative 41; Mismatches 91; Indels 41; Gaps 9;

QY 14 KSAIETADGALDLYNKYLDQVLPWKTFDETIKELSRFKQKYS---QEASVLVGDIKVLLM 70  
DB 438 KPKVEPKDTLAIQDK-ISKI-----ELKNSQLKQLPSQPKESDLNLSNLKVLVS 488  
QY 71 DSODKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKILIRLDGKVKLNEA 130  
DB 489 DDKNKFSELS-----LPQYSISFPKLAGSNNDGTLQVIVQKQKQKVKELK 539  
QY 131 QKSLTSSQSFNNASGKLIALDSQLTNDESEKSYFQSQV---DRIRKAYAGAAAGIVA 187  
DB 540 LTNLLTEFESLKE-----DFQL--DFSNNKORLASSVWMDKIKESLVVKNKTEN 590  
QY 188 GPFGLI-ISYIAAGVIEGKLIPELNNRLKTQVNEFTSLSATVKQ 231  
DB 591 FDFNKYDIDISYSVSS-----LDEVNGKLKIKMTIFKTKDKRLKE 628

RESULT 42

A35186  
salivary agglutinin receptor precursor - Streptococcus sanguis  
C;Species: Streptococcus sanguis  
C;Date: 10-Aug-1990 #sequence\_revision 10-Aug-1990 #text\_change 09-Jul-2004  
C;Accession: A35186  
R;Demuth, D.R.; Golub, E.E.; Malamud, D.  
J. Biol. Chem. 265, 7120-7126, 1990  
A;Title: Streptococcal-host interactions. Structural and functional analysis of a Streptococcus agglutinin receptor precursor.  
A;Reference number: A35186; MUID:90236997; PMID:2185241  
A;Accession: A35186  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1473 <DEM>  
A;Cross-references: UNIPROT:P16952; GB:J05418  
C;Superfamily: surface antigen spa

Query Match 7.0%; Score 106.5; DB 2; Length 1473;

```

Best Local Similarity   23.5%, Pred. No. 60;
Matches      78; Conservative    42; Mismatches      97; Indels 115; Gaps 18;

Qy  15 SALETADGALDLYNKYLDQVIPWKTFTDIKELSRFKQE---YSQBASVLVGDIKVLLMD 71
     :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 123 TATDNAQKQDEIKSDYAKQAEEIKTTTEA-----YKVEAAHQAEATDKINAENKA---- 172
Qy  72 SQDKYPFATQTVVCGVVQTQLLSAYILLDFEYNEKKASAQKOILIRILDGVKKLNEAQ 131
     :||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 173 ADDKYQDLKSHOE---EVEKINTANATAKARYEAELAQQKDLAT-----VKKANE-- 221
Qy 132 KSLILTSSQSFNNAAGKLALDSLTWNDFSSEKSVFSQOVDRIR-----KEAYAGAAGI 185
     ||||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 222 -----DSQ-QDYQNKL SAYOTELARVOKANAENAEKYEKAVKEN 259
Qy 186 VAGPFGLIISYAAGVIEGKLPELNRLK-----TVQNFFTSLSATVKQANKIDA-A 239
     :||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 260 TAKNEAL-----KVENEAIQRNETAKATYEAAMKQYEAADLAA-IKKANEDNDADY 309
Qy 240 KLKLA-----TEIAIGE-----IKTETETTRFYVD-- 265
     ||||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 310 QAKLAAVQOTBLARVOKANAENAEKAYDKAVENTAKNTAIQAENEAIQRNETAKATYDAA 369
Qy 266 ---YD-DLMLSLLKGRAKKMINTCNE--YOOR 291
     :||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 370 VKKYEADL-----AAVRQANATNEADYQAK 394

RESULT 43
Tl3829
Tpr homolog - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: Tl3829
R:Zimowska, G.; Aris, J.P.; Paddy, M.R.
J. Cell Sci. 110, 927-944, 1997
A>Title: A Drosophila Tpr protein homolog is localized both in the extrachromosomal chan
A:Reference number: Z17786; MUID: 97296455; PMID: 9152019
A:Accession: Tl3829
A>Status: preliminary; translated from GR/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2346 <IM>
A:Cross-references: UNIPROT:O01385; EMBL:U91980; NID:g1923273; PID:g1923274; PIDN:AAC475
C:Genetics:
A:Cross-references: FlyBase:FBgn0013756
A:Map position: 2R

```

Db 475 NQENKK 481

RESULT 44  
I41078  
hemolysin - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 09-Jul-2004  
C:Accession: I41078  
R:Schmidt, H.; Beutin, L.; Karch, H.  
Infect. Immun. 63, 1055-1061, 1995  
A:Title: Molecular analysis of the plasmid-encoded hemolysin of Escherichia coli O157:H7  
A:Reference number: I41077; MUID:95172699; PMID:7868227  
A:Accession: I41078  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-998 <RES>  
A:Cross-references: UNIPROT:Q47262; EMBL:X79839; NID:g860924; PIDN:CAA56234.1; PID:g4388;  
C:Superfamily: hemolysin A; hemolysin A homology  
C:Keywords: lipoprotein  
F:233-776/Domain: hemolysin A homology <HLYA>  
P:550,675/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.0%; Score 106; DB 2; Length 998;  
Best Local Similarity 21.6%; Pred. No. 39; Mismatches 117; Indels 66; Gaps 13;  
Matches 66; Conservative 56;

Qy 11 EVKSAIETADGALDLYNKYLQVIPWTKTFOETIKELSRFKQEYSQASVLVGDIKVL-- 68  
Db 165 DIAKSSIELINQVDVTYSSINSTV---DSFSEQLNQLGSLSSKPRLLSSV-GGKLQNLPD 220  
Qy 69 ---LMSQDKYFEATQTVYEWGVVTVQLLSAYILF-----DEYNEKKASAKDKILIRIL 120  
Db 221 LGPLGDLGDL-----VVGILSAVSASFILGNSDAHTGTKAAAGIELTTQVL 266  
Qy 121 DDGVKKLNE---AQK---SLTSSQSFNNAAGKLLALDSQLT-----NDFSEKSSYFQSQV 170  
Db 267 GNVGKAVSQYTLAQRMAQGLSTTAASAGLIITSANMLAISPLSFLAAADKFERAKQLESYS 326  
Qy 171 DRIKEAYAGAAAGIVGAPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFTTSLSATVK 230  
Db 327 ERPKLNYEGDA-----LLAGFKETGAIDA-----GLTTINTVLSVSAGVS 369  
Qy 231 QANK-DIDAAKLLATEATAIGEINTETETTFYVDYDDLMLSLKGAAGKMGINTNEYQ 289  
Db 370 AASSASLIGAPISMLVS-ALGTGISILEASK-----QAMPEHVAEKFAARINEWE 419  
Qy 290 QRHGK 294  
Db 420 KEHGK 424

RESULT 45  
T44967  
gas-vesicle protein gvpC - Halorubrum vacuolatum  
C:Species: Halorubrum vacuolatum  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T44967  
R:Mayr, A.; Pfeifer, F.  
Arch. Microbiol. 168, 24-32, 1997  
A:Title: The characterization of the gvpACNOFGH gene cluster involved in gas vesicle synt  
A:Reference number: 222885; MUID:97361906; PMID:9211710  
A:Accession: T44967  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-540 <MAX>  
A:Cross-references: UNIPROT:O33398; EMBL:Y08571; NID:g2598010; PIDN:CAA69882.1; PID:g258;  
A:Experimental source: strain DSM 3638  
C:Genetics:  
A:Gene: gvpC

Query Match 7.0%; Score 105.5; DB 2; Length 540;  
Best Local Similarity 21.1%; Pred. No. 19;

Qy 288 YQQRHGK 294









**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:42:38 ; Search time 63.2911 Seconds  
(without alignments)  
1717.381 Million cell updates/sec

Title: US-09-993-292B-28

Perfect score: 1508

Sequence: 1 MTSIVADKTVVKNNAIETA.....TCNEYQKHGKTLFEVPEV 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : A\_Geneseq\_23Sep04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID      | Description         |
|------------|-------|---------|--------------|------------|---------------------|
| 1          | 1493  | 99.0    | 309          | 2 AAR67645 | Aar67645 Haemolysin |
| 2          | 1402  | 93.0    | 305          | 6 ABG73265 | Abg73265 Salmonell  |
| 3          | 1402  | 93.0    | 749          | 6 ABG73266 | Abg73266 Salmonell  |
| 4          | 1145  | 75.9    | 1035         | 4 ABG25221 | Abg25221 Novel hum  |
| 5          | 126   | 8.4     | 1095         | 6 ABJ25381 | Abj25381 Aspergill  |
| 6          | 126   | 8.4     | 1277         | 6 ABJ25981 | Abj25981 Aspergill  |
| 7          | 124   | 8.2     | 1036         | 2 Adf07451 | Adf07451 Bacterial  |
| 8          | 120.5 | 8.0     | 962          | 2 AAY31646 | Aay31646 Human tra  |
| 9          | 120.5 | 8.0     | 962          | 6 RAG79733 | Rag79733 Human TAP  |
| 10         | 118   | 7.8     | 627          | 8 ADO52129 | Ado52129 E. coli B  |
| 11         | 118   | 7.8     | 961          | 7 ADB67135 | Adb67135 General v  |
| 12         | 118   | 7.8     | 961          | 8 ADL99406 | Adl99406 Nanosruc   |
| 13         | 118   | 7.8     | 1211         | 5 ABP39975 | Abp39975 Staphyloc  |
| 14         | 117.5 | 7.8     | 476          | 2 AAR43563 | Aar43563 Hyalurona  |
| 15         | 117.5 | 7.8     | 927          | 5 ABB49720 | Abb49720 Listeria   |
| 16         | 116   | 7.7     | 885          | 4 AAG82283 | Aag82283 S. epider  |
| 17         | 116   | 7.7     | 1129         | 7 ADF07734 | Adf07734 Bacterial  |
| 18         | 115.5 | 7.7     | 718          | 8 ADL05067 | Adl05067 M. catarr  |
| 19         | 115   | 7.6     | 2056         | 4 ABB59344 | Abb59344 Drosophil  |
| 20         | 114.5 | 7.6     | 582          | 6 ABM68103 | Abm68103 Photothab  |
| 21         | 114.5 | 7.6     | 1189         | 6 ABU42996 | Abu42996 Protein e  |
| 22         | 114   | 7.6     | 515          | 7 ADC95690 | Adc95690 E. faeciu  |
| 23         | 114   | 7.6     | 796          | 5 ABB92030 | Abb92030 Herbicida  |
| 24         | 113.5 | 7.5     | 1196         | 5 ABB94317 | Abb94317 E. faeciu  |
| 25         | 113   | 7.5     | 458          | 5 ABG80428 | Abg80428 Moraxella  |

|    |       |     |      |            |                    |
|----|-------|-----|------|------------|--------------------|
| 26 | 112.5 | 7.5 | 1020 | 6 ABU30092 | Abu30092 Protein e |
| 27 | 112   | 7.4 | 606  | 2 AAR99673 | Aar99673 Receptor  |
| 28 | 112   | 7.4 | 631  | 2 AAR99675 | Aar99675 RHAMM 1-2 |
| 29 | 112   | 7.4 | 1875 | 6 ABR53560 | AbR53560 Protein s |
| 30 | 112   | 7.4 | 1875 | 7 ADK64380 | Adk64380 Disease t |
| 31 | 111.5 | 7.4 | 808  | 7 ADF04710 | Adf04710 Bacterial |
| 32 | 111   | 7.4 | 435  | 6 ABR57002 | AbR57002 Mouse RHA |
| 33 | 111   | 7.4 | 435  | 6 ABP97484 | Abp97484 Murine RH |
| 34 | 111   | 7.4 | 435  | 7 ADC02471 | Adc02471 Mouse RHA |
| 35 | 111   | 7.4 | 477  | 6 ABR56999 | AbR56999 Mouse RHA |
| 36 | 111   | 7.4 | 477  | 6 ABP97481 | Abp97481 Murine RH |
| 37 | 111   | 7.4 | 477  | 7 ADC02468 | Adc02468 Mouse RHA |
| 38 | 111   | 7.4 | 631  | 5 ABG60843 | Abg60843 Mouse rec |
| 39 | 111   | 7.4 | 794  | 5 AAU11437 | Aau11437 Mouse hya |
| 40 | 111   | 7.4 | 938  | 6 ABU43682 | Abu43682 Protein e |
| 41 | 111   | 7.4 | 1013 | 4 ABB60789 | Abb60789 Drosophil |
| 42 | 111   | 7.4 | 1171 | 5 ABP73682 | Abp73682 Candida a |
| 43 | 110.5 | 7.3 | 304  | 5 ABP26411 | Abp26411 Streptoco |
| 44 | 110.5 | 7.3 | 304  | 8 ADK99510 | Adk99510 Streptoco |
| 45 | 110   | 7.3 | 278  | 8 ADK99511 | Adk99511 Streptoco |
| 46 | 110   | 7.3 | 278  | 8 ADK99512 | Adk99512 Streptoco |
| 47 | 110   | 7.3 | 1163 | 6 ABU23940 | Abu23940 Protein e |
| 48 | 110   | 7.3 | 3187 | 7 ADE56031 | AdE56031 Rat Prote |
| 49 | 110   | 7.3 | 3187 | 7 ADE56035 | AdE56035 Rat Prote |
| 50 | 109.5 | 7.3 | 366  | 3 AAG23551 | Aag23551 Arabidops |

#### ALIGNMENTS

##### RESULT 1

|          |   |
|----------|---|
| AAR67645 |   |
| ID       | AAR67645 standard; protein; 309 AA.                                       |
| XX       |   |
| AC       | AAR67645;   |
| XX       |   |
| DT       | 25-MAR-2003 (revised)   |
| DT       | 06-JUL-1995 (first entry)   |
| XX       |   |
| DE       | Haemolysin regulator.   |
| XX       |   |
| KW       | Haemolysin regulator; tuberculosis; vaccine; therapy.                     |
| XX       |   |
| OS       | Escherichia coli.   |
| XX       |   |
| PN       | WO9428137-A1.   |
| XX       |   |
| PD       | 08-DEC-1994.  |
| XX       |   |
| PF       | 24-MAY-1994; 94WO-US005869.   |
| XX       |   |
| PR       | 24-MAY-1993; 93US-00066830.   |
| XX       |   |
| PA       | (USSH ) US DEPT HEALTH & HUMAN SERVICES.                                  |
| XX       |   |
| PI       | King CH, Shinnick TM, Sathish M;  |
| XX       |   |
| DR       | WPI; 1995-022809/03.  |
| XX       |   |
| DR       | P-PSDB; AAQ75857.   |
| XX       |   |
| PT       | New nucleic acid encoding haemolytic regulator of E. coli - and derived   |
| XX       |   |
| PT       | vectors, proteins etc used to transform Mycobacterium cells to produce    |
| XX       |   |
| PS       | antituberculosis vaccines providing improved immune response.             |
| XX       |   |
| CC       | Claim 9; Page 39-40; 46pp; English.                                       |
| XX       |   |
| CC       | The sequence represents the E. coli haemolysin regulator which may be     |
| XX       |   |
| CC       | used as an immunogen in a vaccine composition against tuberculosis        |
| XX       |   |
| CC       | (Mycobacterium tuberculosis). (Updated on 25-MAR-2003 to correct PN       |
| XX       |   |
| CC       | field.) (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR- |
| XX       |   |
| CC       | 2003 to correct PA field.)  |
| XX       |   |
| SQ       | Sequence 309 AA;  |

Query Match 99.0%; Score 1493; DB 2; Length 309;  
Best Local Similarity 99.7%; Pred. No. 5.3e-118;  
Matches 300; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQESQAASV 60  
DB 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQESQAASV 60  
QY 61 LVGDIKTLMDSDQKYPEATQTVVWCGVATQLLAAVILLDFEYNEKKAQKIDILIKVL 120  
DB 61 LVGDIKTLMDSDQKYPEATQTVVWCGVATQLLAAVILLDFEYNEKKAQKIDILIKVL 120  
QY 121 DDGITKLINEAKSLVSSQFNNSAGKLALDLSQTLNDFSEKSYFOSQVDKIRKEAYAG 180  
DB 121 DDGITKLINEAKSLVSSQFNNSAGKLALDLSQTLNDFSEKSYFOSQVDKIRKEAYAG 180  
QY 181 AAAGVVAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLSNTVKQANKDIDAAK 240  
DB 181 AAAGVVAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLSNTVKQANKDIDAAK 240  
QY 241 LKLTTEIAAIGEIKTETETTRFYVDYDDLMLSLKKAQKMTNCNEYQKHGKKTLEFV 300  
DB 241 LKLTTEIAAIGEIKTETETTRFYVDYDDLMLSLKKAQKMTNCNEYQKHGKKTLEFV 300  
QY 301 P 301  
DB 301 P 301

RESULT 2  
ABG73265  
ID ABG73265 standard; protein; 305 AA.  
XX AC ABG73265;  
XX 17-APR-2003 (first entry)  
XX Salmonella typhi cytolysin A (ClyA) protein.  
XX Protein expression vector; expression cassette; export protein;  
KW transformed host cell; bacterial cell; immune response; HlyE family;  
KW cytolysin A; ClyA; recombinant protein; bioreactor; bacterial growth;  
KW live vector immunogenic composition; animal feed; animal vaccination;  
KW food industry; nutritional supplement; biomediation; waste disposal;  
KW waste treatment; host bacterium.  
XX Salmonella typhi.  
OS US2002146430-A1.  
XX PN 10-OCT-2002.  
XX PD 23-NOV-2001; 2001US-00993292.  
XX PF 22-NOV-2000; 2000US-0252516P.  
XX PR (GALE/) GALEN J E.  
XX PA Galen JE;  
XX PI WPI; 2003-228013/22.  
XX DR N-PSDB; ABX15174.  
XX DR Novel system for expressing protein of interest, has expression vector  
PT with cassette having export protein coding sequence fused to protein  
PT coding sequence, host cell transformed with vector, and culturing  
PT environment.  
XX Claim 5; Page 18; 33pp; English.  
XX PS The present invention relates to a system for expressing a protein of  
XX interest. The system comprises an expression vector having an expression  
CC

CC cassette, where the expression cassette comprises an export protein  
CC coding sequence genetically fused to a DNA sequence encoding the protein  
CC of interest, a host cell transformed with the expression vector, and a  
CC culturing environment for the transformed host cell, where the expression  
CC cassette expresses the export protein-protein of interest fusion protein  
CC which is exported out of the host cell. The system is useful for  
CC expressing a gene in a bacterial cell, by providing an expression vector  
CC to a population of untransformed bacterial cells, and expressing the  
CC fusion protein is produced and exported into the culture medium. The  
CC expression cassette, such that the export protein-protein of interest  
CC protein of interest is preferably an antigen and this method is useful  
CC for eliciting an immune response from a host against the fusion protein.  
CC The export protein may be a member of the HlyE family such as cytolysin A  
CC (ClyA). The system is useful for the production of recombinant proteins  
CC inside a bacterial host cell, in a bioreactor, and various devices that  
CC facilitate the growth of bacteria. The system is also useful for  
CC providing an animal antigenic material against which an immune response  
CC may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can  
CC be used as a live vector immunogenic composition capable of facilitating  
CC the generation of an immune response in an animal. The system is also  
CC useful for preparing immunogenic compositions for vaccinating animals,  
CC and for use in the food, nutritional supplement, animal feed,  
CC biomediation, waste disposal, and waste treatment industries. The system  
CC is very advantageous for protein expression, as proteins that are toxic  
CC to host bacterium can also be expressed. A population of recombinant host  
CC cells can be transformed with the expression vector, and then the  
CC population of bacterial host cells can be maintained in culture and used  
CC to produce proteins without having to harvest and lyse the host cells.  
CC The present represents Salmonella typhi ClyA protein  
XX  
XX Sequence 305 AA;

Query Match 93.0%; Score 1402; DB 6; Length 305;  
Best Local Similarity 91.1%; Pred. No. 2.7e-110;  
Matches 276; Conservative 16; Mismatches 11; Indels 0; Gaps 0;  
QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQESQAASV 60  
DB 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQESQAASV 60  
QY 61 LVGDIKTLMDSDQKYPEATQTVVWCGVATQLLAAVILLDFEYNEKKAQKIDILIKVL 120  
DB 61 LVGDIKTLMDSDQKYPEATQTVVWCGVATQLLAAVILLDFEYNEKKAQKIDILIKVL 120  
QY 121 DDGITKLINEAKSLVSSQFNNSAGKLALDLSQTLNDFSEKSYFOSQVDKIRKEAYAG 180  
DB 121 DDGITKLINEAKSLVSSQFNNSAGKLALDLSQTLNDFSEKSYFOSQVDKIRKEAYAG 180  
QY 181 AAAGVVAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLSNTVKQANKDIDAAK 240  
DB 181 AAAGVVAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLSNTVKQANKDIDAAK 240  
QY 241 LKLTTEIAAIGEIKTETETTRFYVDYDDLMLSLKKAQKMTNCNEYQKHGKKTLEFV 300  
DB 241 LKLTTEIAAIGEIKTETETTRFYVDYDDLMLSLKKAQKMTNCNEYQKHGKKTLEFV 300  
QY 301 PEV 303  
DB 301 PDV 303

RESULT 3  
ABG73266  
ID ABG73266 standard; protein; 749 AA.  
XX AC ABG73266;  
XX 17-APR-2003 (first entry)  
XX Salmonella typhi ClyA-Bacillus subtilis sacB fusion protein.  
XX Protein expression vector; expression cassette; export protein;  
KW transformed host cell; bacterial cell; immune response; HlyE family;  
KW



CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activities. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1035 AA;  
Query Match 75.9%; Score 1145; DB 4; Length 1035;  
Best Local Similarity 100.0%; Pred. No. 8.4e-88;  
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 70 MDSQKYFEATQVYEWGCVATQLLAAYILLFDEYNEKKAQAQDILIKVLDGDTKLE 129  
DB 1 MDSQKYFEATQVYEWGCVATQLLAAYILLFDEYNEKKAQAQDILIKVLDGDTKLE 60  
QY 130 AKSLVSSQSFNNASGKLLALDSQTLNDSEKSSYFQSDVKIRKEAYAGAAAGVWAGP 189  
DB 61 AQKSLVSSQSFNNASGKLLALDSQTLNDSEKSSYFQSDVKIRKEAYAGAAAGVWAGP 120  
QY 190 FGLIISYSIAAGVWEGKLIPELKNKLSQVQFFTLSTNTVQKAKDIDAAKLUKTTTEIAA 249  
DB 121 FGLIISYSIAAGVWEGKLIPELKNKLSQVQFFTLSTNTVQKAKDIDAAKLUKTTTEIAA 180  
QY 250 IGEIKTETTRFYDDYDMLSLKEAKKQKNTCNEYQKHGKTLFE 299  
DB 181 IGEIKTETTRFYDDYDMLSLKEAKKQKNTCNEYQKHGKTLFE 230  
RESULT 5  
ABU25381  
ID ABU25381 standard; protein; 1095 AA.  
XX  
AC ABU25381;  
XX  
DT 16-APR-2003 (first entry)  
XX  
DE Aspergillus fumigatus essential gene protein #39.  
XX  
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;  
KW cancer; contamination; biofilm; antibody; immune response.  
XX  
OS Aspergillus fumigatus.  
XX  
FN WO200286090-A2.  
XX  
PD 31-OCT-2002.  
XX  
PF 23-APR-2002; 2002WO-US013142.  
XX  
PR 23-APR-2001; 2001US-0285697P.  
PR 27-APR-2001; 2001US-0287066P.  
PR 05-JUN-2001; 2001US-0295890P.  
PR 09-JUL-2001; 2001US-0303899P.  
PR 31-AUG-2001; 2001US-0316362P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Jiang B, Tishkoff D, Zamudio C, Eroehkin AM, Hu W, Lemieux SM;  
XX  
DR WPI; 2003-093124/08.  
XX  
PT New purified or isolated nucleic acids of essential genes of Aspergillus  
PT fumigatus, useful for treating or preventing infections by A. fumigatus,  
PT or for treating a non-infectious disease in a subject e.g. cancer.  
XX

PS Disclosure; Page; 175pp; English.  
XX  
CC The invention relates to novel purified or isolated nucleic acids of  
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of  
CC the invention are used to treat or prevent infections by a pathogenic  
CC organism such as A. fumigatus, to treat a non-infectious disease in a  
CC subject (e.g. cancer), to prevent or inhibit formation of an object  
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a  
CC biofilm comprising A. fumigatus. The polynucleotides are useful for  
CC expressing recombinant protein for characterization, screening or  
CC therapeutic use, as markers for host tissues in which the pathogenic  
CC organisms invade or reside, for comparing with the DNA sequence of A.  
CC fumigatus to identify duplicated genes or paralogues having the same or  
CC similar biochemical activity and/or function, for comparing with DNA  
CC sequences of other related or distant pathogenic organisms to identify  
CC potential orthologous essential or virulence genes, for selecting and  
CC making oligomers for attachment to a nucleic acid array for examination  
CC of expression patterns, for raising anti-protein antibodies, as an  
CC antigen to raise anti-DNA antibodies or to elicit another immune  
CC response, and for identifying polynucleotides encoding the other protein  
CC with which binding occurs or to identify inhibitors of the binding  
CC interaction. The polypeptides may be used to raise antibodies or to  
CC elicit immune response, as a reagent in assays designed to quantitatively  
CC determine levels of the protein in biological fluids, as a marker for  
CC host tissues in which pathogenic organism invade or reside, and to  
CC isolate correlative receptors or ligands in the case of virulence  
CC factors. This sequence represents a protein of one of the essential genes  
CC of Aspergillus fumigatus of the invention  
XX  
SQ Sequence 1095 AA;  
Query Match 8.4%; Score 126; DB 6; Length 1095;  
Best Local Similarity 21.4%; Pred. No. 0.2;  
Matches 73; Conservative 49; Mismatches 73; Indels 146; Gaps 15;  
QY 4 IVADKTVVVKNAIETADGALDLYNKYLDQVFPWQTFDETIKELGRFQKQYSAASV--L 61  
DB 281 IKKEKIDIEATNALVPFDEKVDITRKVER-----FASRIAEG--KERDQAANVKOL 332  
QY 62 VGDITKLLMDSQDKYFEATQVYEWGCVATQLLAAYILLFDEYNEKKAQAQDILIKVLD 121  
DB 333 EKDLK--VVEKAQAQWEA-----EW----- 350  
QY 122 DGITKLENAQKSLVSSQSFNNASGKLLALDSQ---LTNDFSEKSSYFQSDVKIRKEA 177  
DB 351 -----HKTNSKGGQLSESDQOEYRMLKEEYKRSSEAEQINLDNLKROR 394  
QY 178 YAGAAAGVWAGPFLIISYSIAAGVWEGKLIPELKNK-----LKSQVQFFTT----- 224  
DB 395 KTEAEA-----YN-----SLKSKFDSQTEWQLKSVENDTQTLTERKS 430  
QY 225 -LSNTVQKAKDIDAAKLEK---LTTEIAAIGEIKTETTRFYDDYDMLSLKEA--- 277  
DB 431 ALNDTVKTTSKEDIRKKKELNALTSELRISQWTELE-----EKQVVLKLEADGG 484  
QY 278 -----AKKMINT-----CNEYQKHOK 294  
DB 485 KQTERELRAKELISTLKRIPPGVKGVRSDLCRPFQKKYAE 525  
RESULT 6  
ABU25981  
ID ABU25981 standard; protein; 1277 AA.  
XX  
AC ABU25981;  
XX  
DT 16-APR-2003 (first entry)  
XX  
DE Aspergillus fumigatus essential gene protein #639.  
XX  
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;  
KW cancer; contamination; biofilm; antibody; immune response.  
XX

OS Aspergillus fumigatus.  
 PN WO200286090-A2.  
 XX 31-OCT-2002.  
 XX 23-APR-2002; 2002WO-US013142.  
 XX 23-APR-2001; 2001US-0285697P.  
 PR 27-APR-2001; 2001US-0287066P.  
 PR 05-JUN-2001; 2001US-0295890P.  
 PR 09-JUL-2001; 2001US-0303899P.  
 PR 31-AUG-2001; 2001US-0316362P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX JIANG B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;  
 XX WPI; 2003-093124/08.  
 XX New purified or isolated nucleic acids of essential genes of *Aspergillus*  
 PT fumigatus, useful for treating or preventing infections by *A. fumigatus*,  
 PT or for treating a non-infectious disease in a subject e.g. cancer.  
 XX Disclosure; Page; 175pp; English.  
 XX The invention relates to novel purified or isolated nucleic acids of  
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of  
 CC the invention are used to treat or prevent infections by a pathogenic  
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a  
 CC subject (e.g. cancer), to prevent or contain contamination of an object  
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a  
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for  
 CC expressing recombinant protein for characterisation, screening or  
 CC therapeutic use, as markers for host tissues in which the pathogenic  
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*  
 CC *fumigatus* to identify duplicated genes or paralogues having the same or  
 CC similar biochemical activity and/or function, for comparing with DNA  
 CC sequences of other related or distant pathogenic organisms to identify  
 CC potential orthologous essential or virulence genes, for selecting and  
 CC making oligomers for attachment to a nucleic acid array for examination  
 CC of expression patterns, for raising anti-protein antibodies, as an  
 CC antigen to raise anti-DNA antibodies or to elicit another immune  
 CC response, and for identifying polynucleotides encoding the other protein  
 CC with which binding occurs or to identify inhibitors of the binding  
 CC interaction. The polypeptides may be used to raise antibodies or to  
 CC elicit immune response, as a reagent in assays designed to quantitatively  
 CC determine levels of the protein in biological fluids, as a marker for  
 CC host tissues in which pathogenic organism invade or reside, and to  
 CC isolate correlative receptors or ligands in the case of virulence  
 CC factors. This sequence represents a protein of one of the essential genes  
 CC of *Aspergillus fumigatus* of the invention  
 XX SQ Sequence 1277 AA;  
 Query Match 8.4%; Score 126; DB 6; Length 1277;  
 Best Local Similarity 21.4%; Pred. No. 0.24;  
 Matches 73; Conservative 49; Mismatches 73; Indels 146; Gaps 15;  
 4 IVADKTVVVKNAIETADGALDLYNKYLDVPIPWQTFDETIELSRFKQVYQASV--L 61  
 323 IKKEDLEENALVPVDEKVDITRKVER-----FASRIAETG--KERDQAANYKQL 374  
 62 VGDITKLMSDQKYFEATQVYEWCGVATOLLAAAYILLFDEYNKKASAKOILIKVLD 121  
 375 EKDLK--VWEKAQAQWEA-----EW-----  
 122 DGITKLNEAKSLIVSOSFNAGSKLALDSQ-----LTNDFSKSSYFQSQVDKIRKEA 177  
 393 -----HKTMSNGGQLSESQDQYKMLKEVRSSAEQINLDNLKQR 436  
 178 YAGAAAGVAGPFGILISYSTAAGVGEGLTPELKNK-----LKSQNEFTT----- 224

437 KTEAEA-----YN-----SLKSKFSDSTEWQLKSVENDTQTLTERKS 472  
 225 -LSNTVKQANKOIDAANKL---LTTEIAAIGEIKTETETTRFYVDYDMLSLKEA--- 277  
 473 ALNDTVKTSKEIDRKKKELNALTSERLRISQMRTELE-----EKVQVVLKLLLEADDG 526  
 278 -----AKMINT-----CNEYQKRHGK 294  
 527 KKQTERELRAKELISTLKRIFPGVGRVSDLCRPKQKKYAE 567  
 RESULT 7  
 ADF07451  
 ID ADF07451 standard; protein; 1036 AA.  
 AC ADF07451;  
 XX 12-FEB-2004 (first entry)  
 DT Bacterial polypeptide #3564.  
 DE *Proteus mirabilis* infection; bacterial infection; antibacterial;  
 XX immunostimulant.  
 KW *Proteus mirabilis*.  
 XX *Proteus mirabilis*.  
 OS US6605709-B1.  
 PN 12-AUG-2003.  
 XX 05-APR-2000; 2000US-00543681.  
 PF 09-APR-1999; 99US-0128706P.  
 PR (GENO-) GENOME THERAPEUTICS CORP.  
 PA Breton GL;  
 PI WPI; 2003-895291/82.  
 DR N-PSDB; ADF03279.  
 XX New *Proteus mirabilis* polypeptides and polynucleotides, useful as  
 PT reagents for diagnosis of bacterial disease, as components of  
 PT antibacterial vaccines, as targets for antibacterial drugs, or as  
 PT biocontrol agents for plants.  
 XX Disclosure; SEQ ID NO 7736; 870pp; English.  
 XX The invention relates to new *Proteus mirabilis* polypeptides and  
 CC polynucleotides. The invention also relates to antibodies against the  
 CC polypeptides, methods for producing the polypeptides, a method of  
 CC generating vaccines for immunising an individual against *P. mirabilis*, a  
 CC method for evaluating a compound for the ability to bind a *P. mirabilis*  
 CC polypeptide and a method for screening test compounds for anti-bacterial  
 CC activity. The polypeptides and polynucleotides are useful as molecular  
 CC targets for diagnosing, preventing and treating pathological conditions  
 CC resulting from bacterial infection, as reagents for diagnosis of  
 CC bacterial diseases, as components of antibacterial vaccines, as targets  
 CC for antibacterial drugs or as bio-control agents for plants. This  
 CC sequence represents a *Proteus mirabilis* polypeptide of the invention.  
 XX SQ Sequence 1036 AA;  
 Query Match 8.2%; Score 124; DB 7; Length 1036;  
 Best Local Similarity 21.5%; Pred. No. 0.27;  
 Matches 82; Conservative 56; Mismatches 127; Indels 116; Gaps 16;  
 4 IVADKTVVVKNAIETADGALDLYNKYLDVPIPWQTFDETIELSRFKQVYQASV--L 61  
 178 IKKQKNGENVSSS-ELAKASIDLINQLVD-----TASSLNNNNISAFSQNLKGLSVLSN 230  
 62 -----VGDITKLMSDQKYFEATQVYEWCGVATOLLAAAYILL-FDEYNKKASAKO 114

Db 231 TKHLNGVN-KLONPLNLDKLTGLDVT-...SGILSAISAFILSNADADGTGAAGVE 286  
QY 115 ILIKVLD-...GITKLEAQKSLVSSQSFNNASGKLLALDSQL--T 156  
Db 287 LTTKVLGVNGKAVSQYILAQVAGLS- TSAASAGLIAVTLAISPLSLATADQFKRA 345  
QY 157 NDFSEKSYFQ-...SOVDKIRKEA-...YAGAAAGVVAGPFG 192  
Db 346 NKIEYSQRFKFGVGDGSLAFAFRKETGAIDASLTINTALGTITISAGISAASLASLIGA 405  
QY 193 IISVIAA--GVVEKLIPKLNKLVKSVQNFVTLSTNV-... 229  
Db 406 FIALVGAITGIISGLEASQSMFEHVAN---RWANTIAEWKTHGNFENGYDARHS 462  
QY 230 -----KQANKDIDAALKLIT-...EIAAGEIKTETETTRFFVYDIDL 270  
Db 463 AFLEDNFKLLSQYNKEYSVERSVLITQOHWDDELIGELASVTNGAKTILSGKSYIDY---- 518  
QY 271 LSLKKAAMKMTNCEYQKR 291  
Db 519 ----YEEGRLEKKEPQK 535

RESULT 8

AA31646  
ID AAY31646 standard; protein; 962 AA.

AC AAY31646;

DT 02-NOV-1999 (first entry)

XX Human transport-associated protein-8 (TRANP-8).

DE Transport-associated protein; TRANP; nuclear pore; nuclear transport;  
KW vesicle trafficking; cancer; cystic fibrosis; multidrug resistance;  
KW hypercholesterolaemia; diagnosis; treatment.

OS Homo sapiens.

| Key               | Location/Qualifiers                           |
|-------------------|---|
| Modified-site 18  | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 34  | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 74  | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 81  | /note= "O-phosphorylated by tyrosine kinase"  |
| Modified-site 91  | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 101 | /note= "N-glycosylated"                       |
| Modified-site 123 | /note= "N-glycosylated"                       |
| Modified-site 129 | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 243 | /note= "N-glycosylated"                       |
| Modified-site 336 | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 410 | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 451 | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 453 | /note= "N-glycosylated"                       |
| Modified-site 585 | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 631 | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 632 | /note= "O-phosphorylated by casein kinase II" |
| Modified-site 717 | /note= "O-phosphorylated by casein kinase II" |

|    |               |   |
|----|---------------|---|
| FT | Modified-site | /note= "O-phosphorylated by casein kinase II" |
| FT | Modified-site | /note= "O-phosphorylated by casein kinase II" |
| FT | Modified-site | /note= "O-phosphorylated by casein kinase II" |
| FT | Modified-site | /note= "O-phosphorylated by casein kinase II" |
| FT | Modified-site | /note= "O-phosphorylated by casein kinase II" |
| FT | Modified-site | /note= "N-glycosylated"                       |
| FT | Modified-site | /note= "O-phosphorylated by casein kinase II" |
| FT | Modified-site | /note= "O-phosphorylated by casein kinase II" |

WO9941373-A2.

19-AUG-1999.

05-FEB-1999; 99WO-US002527.

11-FEB-1998; 98US-00021764.

(INCY-) INCYTE PHARM INC.

Au-Young J, Hillman JL, Lal P, Guegler KJ, Corley NC, Yue H;  
Bandman O, Baughn MR;

WPI; 1999-508646/42.

N-PSDB; AAZ11738.

Human TRANP coding sequences, used to treat transport disorders and cancer.

Claim 1; Page 74-77; 87pp; English.

This sequence represents human transport-associated protein-8 (TRANP-8). The DNA sequence was first identified in a human colon tissue cDNA library. The full-length cDNA was derived from a series of overlapping and/or extended cDNA sequences and is a consensus. TRANP-1 to 9 (AAV31639 -V31647) are a novel group of proteins with chemical and structural homology that are involved in the transport of molecules, either intracellularly or to the extracellular environment. Examples of such disorders include cystic fibrosis, multidrug resistance, hypercholesterolaemia and certain forms of diabetes mellitus. Defective nuclear transport may play a role in cancer. For example, the BRCA1 protein, associated with familial breast cancer, is normally located in the nucleus via nuclear pore complexes, but is aberrantly located in the cytoplasm in breast cancer cells. In other cancers, cells can secrete excessive amounts of hormones e.g. cancers of the adrenal medulla can secrete excessive amounts of adrenaline and noradrenaline, leading to hypertension. TRANP is expressed in cancer cells, and transport disorders result from either excessive or insufficient molecular transport. Anti-TRANP antibodies and nucleic acids encoding TRANP can be used as diagnostic tools for such disorders. TRANP antagonists can be used to treat or prevent a cancer associated with increased TRANP expression. Anti-TRANP antibodies can be used directly as an antagonist or as a targeting mechanism for drugs. Alternatively, a TRANP antisense nucleotide can be used to treat cancers. A TRANP agonist or expression vector may be used to treat a disorder caused by reduced transport of biologically active molecules

Sequence 962 AA;

Query Match 8.0%; Score 120.5; DB 2; Length 962;

Best Local Similarity 23.7%; Pred No. 0.48;

Matches 74; Conservative 51; Mismatches 136; Indels 51; Gaps 13;

QY 7 DKTVEVKNIAETADGALDLY-NKYLDQVWPQTFDETIKLSRFKQYSQAASVLVCDI 65

635 DKKEEVKTLQHDNIVTHYKNMIREQDLQLELRQVSTLKQNEQLQTAVTQVQSQI 694

Db



Qy 66 K-----TLIM-----DSQDKYFEATQTVYECGVAATQLLAAAYILLFDEYNKKAS-- 110  
 Db 695 QQHKDQYNLLKIQLGKDNQHGSGYSEGAQ-----MNGIQPBEIGR---LREIEELKRNQE 747  
 Qy 111 -----AQKDILIKVLDGDIKLENAQKSLIVS---SQSFNNAAGKLLALDSQLTNDPSE 161  
 Db 748 LLQSLQTEKSMIENKSSQTSCTNEQSSAIVSARDSEQVAELKQELATLKSQ--NQSV 806  
 Qy 162 KSSYFQSQVDKIRKEAYAGAAAGVAGPFGHIIISYSIAAGVVEGKL-----IPELKNKL 215  
 Db 807 EITKLQTEKQELLQKTEAFKSVQGBETETIIATKTDD--VEGRSALLQETKELKNEI 864  
 Qy 216 KSVQNFPTLISNTVKQANKDIDAAKLKTTEIAAIGELKTETETTRFVYDDMLSLK 275  
 Db 865 KALSEERTAIKEQLDSSNSTIAI-----LQTE-----KDKLEITDSKKEQDDLLV-LLA 914  
 Qy 276 EAAKMINTCNE 287  
 Db 915 DQDQKILSLKNK 926

## RESULT 9

AAG79733  
 ID AAG79733 standard; protein; 962 AA.

AC AAG79733;

DT 04-MAR-2003 (first entry)

DE Human TAP.

XX Enzyme; human; insulin-responsive aminopeptidase; IRAP;  
 KW insulin response modulator; transcytosis-associated protein; TAP;  
 KW GLUT4-containing vesicle; insulin; translocation; glucose clearance;  
 KW blood glucose; sensitivity; resistance; diabetes; prediabetes;  
 KW hyperglycemia; Type I diabetes.

XX Homo sapiens.

XX WO200209032-A2.

XX 14-NOV-2002.

XX 08-MAY-2002; 2002WO-US014493.

XX 08-MAY-2001; 2001US-0289669P.

XX (ADIP-) ADIPOGENIX INC.

XX (TAKE ) TAKEDA CHEM IND LTD.

XX Brooks CC;

XX WPI; 2003-112000/10.

XX Identifying an insulin response modulator comprises contacting a  
 PT composition comprising insulin-responsive aminopeptidase (IRAP) and  
 PT transcytosis-associated protein (TAP) with test compound and determining  
 PT modulation of IRAP:TAP binding.

XX Disclosure; Page 47-49; 51pp; English.

XX This sequence represents human transcytosis-associated protein (TAP). This  
 CC sequence may be used in the method of the invention for identifying an  
 CC insulin response modulator. The method comprises contacting a composition  
 CC comprising insulin-responsive aminopeptidase (IRAP), or its fragment, and  
 CC TAP, or its fragment, with a test compound and determining the ability of  
 CC the test compound to modulate binding of IRAP, or its fragment, to TAP,  
 CC or its fragment. TAP interacts with IRAP in GLUT4 -containing vesicles.  
 CC Binding of TAP to IRAP is competitively inhibited by a bioactive fragment  
 CC comprising residues 1-109 of IRAP. TAP is thought to function as a tether  
 CC which secures GLUT4 vesicles to an internal cellular component. The  
 CC method of the invention is useful for identifying an insulin response

CC modulator, especially an IRAP:TAP modulator. The modulator is useful for  
 CC modulating GLUT4 translocation, enhancing glucose clearance in an insulin  
 CC resistant subject, and regulating blood glucose levels in a subject.  
 CC Furthermore, The modulator is useful for treating a subject having an  
 CC insulin response disorder, e.g. reduced insulin sensitivity or insulin  
 CC resistance, or diabetes, for treating a subject having prediabetes or its  
 CC symptoms, e.g. hyperglycemia and/or Type I diabetes

XX Sequence 962 AA;

Query Match 8.0%; Score 120.5; DB 6; Length 962;  
 Best Local Similarity 23.7%; Pred. No. 0.48;  
 Matches 74; Conservative 51; Mismatches 136; Indels 51; Gaps 13;

Qy 7 DKTVEVVKNIAIETADGALDLY-NKLDQVTPMQTFDETIKELSRPKQBYSOAAASVVLGDI 65  
 Db 635 DKKEEVKKTLEQHDNIVTHYKNMIREQDLQLEBLRQQVSTLKQNEQLQTAVTQQVSOI 694  
 Qy 66 K-----TLIM-----DSQDKYFEATQTVYECGVAATQLLAAAYILLFDEYNKKAS-- 110  
 Db 695 QQHKDQYNLLKIQLGKDNQHGSGYSEGAQ-----MNGIQPBEIGR---LREIEELKRNQE 747  
 Qy 111 -----AQKDILIKVLDGDIKLENAQKSLIVS---SQSFNNAAGKLLALDSQLTNDPSE 161  
 Db 748 LLQSLQTEKSMIENKSSQTSCTNEQSSAIVSARDSEQVAELKQELATLKSQ--NQSV 806  
 Qy 162 KSSYFQSQVDKIRKEAYAGAAAGVAGPFGHIIISYSIAAGVVEGKL-----IPELKNKL 215  
 Db 807 EITKLQTEKQELLQKTEAFKSVQGBETETIIATKTDD--VEGRSALLQETKELKNEI 864  
 Qy 216 KSVQNFPTLISNTVKQANKDIDAAKLKTTEIAAIGELKTETETTRFVYDDMLSLK 275  
 Db 865 KALSEERTAIKEQLDSSNSTIAI-----LQTE-----KDKLEITDSKKEQDDLLV-LLA 914  
 Qy 276 EAAKMINTCNE 287  
 Db 915 DQDQKILSLKNK 926

## RESULT 10

ADOS2129  
 ID ADOS2129 standard; protein; 627 AA.

XX AC ADOS2129;

XX DT 12-AUG-2004 (first entry)

XX E. coli Bla SS-pspAEF5668-pspARx1 chimeric protein, alternative version.

XX Immunogenic composition; vaccine; Th2-type immune response;  
 KW pneumococcal surface protein A; PspA; beta-lactamase; bla; chimeric.

XX Streptococcus pneumoniae.

XX Escherichia coli.

XX Chimeric.

XX Key Location/Qualifiers

XX FH Region 1..37

XX FT /note= "Escherichia coli beta-lactamase signal peptide  
 (bla SS)"

XX FT Region 38..269

XX FT /note= "Streptococcus pneumoniae pneumococcal surface  
 protein A (PspA)EF5668"

XX FT Region 270..627

XX FT /note= "Streptococcus pneumoniae pneumococcal surface  
 protein A (PspA)Rx1"

XX FT Misc-difference 624..625

XX /note= "Encoded by bases GAATAATAGAAG"

XX US2004101531-A1.

XX PD 27-MAY-2004.

XX

15-APR-2003; 2003US-00414532.  
 16-APR-2002; 2002US-0372710P.  
 (CURT/) CURTISS R.  
 (KANG/) KANG H Y.  
 Curtiss R, Kang HY;  
 WPI; 2004-399655/37.  
 N-PSDB; ADO52074.  
 New vaccine comprising a live attenuated strain of pathogenic gram-negative bacteria, useful in eliciting a Th2-type immune response in a vertebrate against pathogens, e.g., helminths, fungi, viruses, protozoans or bacteria.  
 Example 5; Fig 13; 94pp; English.  
 The invention relates to immunogenic compositions and vaccines comprising a live attenuated strain of pathogenic gram negative bacteria that secrete an antigen. The vaccine is useful in eliciting a Th2-type immune response in a vertebrate against pathogens, e.g., helminths, fungi, viruses, protozoans or bacteria. The present sequence is Escherichia coli beta-lactamase signal peptide (bla SS)-Streptococcus pneumoniae pneumococcal surface protein A (PspA)EF5668-ppARX1 chimeric protein. This sequence is used in the exemplification of the invention. Note: This sequence is stated to be similar to the sequence shown in the sequence listing, however these sequences differ.  
 Sequence 627 AA;  
 Query Match 7.8%; Score 118; DB 8; Length 627;  
 Best Local Similarity 22.5%; Pred. No. 0.44; Mismatches 128; Indels 94; Gaps 15;  
 Matches 79; Conservative 50;  
 QY 11 EVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKOYSQAASVVLGDIKTLIM 70  
 Db 173 DYKEGLEQA-----ITNKAELATTQNIIDTKQDLEAELEBKVLATLDPGKT--Q 225  
 QY 71 DSQDKVFEATQVYVCGVATQLLAAYILLFDEYNEKASAKDILIKVLDDGITKL--- 127  
 Db 226 DELDK--EAEAE-----ELNEKVEALQNO--VAELEELSKLEDN 261  
 QY 128 -----NEAKSLVSSQSFNNASGKL-----LALD-----SQTNDPFEKSSY 165  
 Db 262 LKDAETLQSPVASQSKAKDYDAKKDAKNAKKADEDAQKALDDAKAAQKYDEQKTE 321  
 QY 166 FQSQVDKIRKEAVAGAAAGVAGPFGLIISYSTAAGVVEGKLIPELKNKLSQVNFPTT- 224  
 Db 322 EKALEKAASEMDKAVAAVQQAAYLAYQQAATDKAAKDAADKMIDEAKGEEEAKTFTV 381  
 QY 225 -----LSNTVQA-----NKDIDAAKLKLTETIAAIGEIKTE-----TE 258  
 Db 382 RAMVVPPEQLAETKKSEAKQAPELTKLSEAKKL--EAAALEKAASEMDKAVAA 440  
 QY 259 TTRFVVDYDILMLSLLEAKAKMINTCNEYQRHGK-KTLFE-----VPE 302  
 Db 441 VQQAAYLAYQQAATDKAAKDAADKMI---DEAKGEEEAKTFTVRAVMPVE 488  
 RESULT 11  
 ID ADB67135  
 ID ADB67135 standard; protein; 961 AA.  
 XX AC ADB67135;  
 XX AC ADB67135;  
 XX AC ADB67135;  
 DT 04-DEC-2003 (first entry)  
 DE General vesicular transport factor p115 SEQ ID NO:155.  
 DE staged assembly; nanostructure; peptide nucleic acid; PNA;  
 KW structural reinforcement; aerogel; paper; plastic; cement;  
 KW tensile strength; identification marker; anti-counterfeiting marker;  
 KW enzyme support; catalyst support; assembly scaffold; nanowire;  
 KW nanocircuit; molecular sieve; molecular filter; biosensor.  
 OS Bos taurus.  
 XX WO2003072829-A1.  
 XX PD 04-SEP-2003.  
 XX 21-FEB-2003; 2003WO-US005390.  
 XX 21-FEB-2002; 2002US-00080608.  
 PA (NANO-) NANOFAMES INC.  
 XX Hyman PL, Goldberg EB;  
 XX WPI; 2003-721788/68.  
 Staged assembly of nanostructures, useful e.g. in biosensors or as catalyst supports, using assembly units derived from peptide nucleic acids.  
 Disclosure; Page 58; 118pp; English.  
 The present invention describes a method (M1) for the staged assembly of a nanostructure using peptide nucleic acids (PNAs). M1 comprises: (a) contacting a nanostructure intermediate (NSI) having at least one unbound joining element (JE) with an assembly unit (AU) that comprises several different JE where: (i) none of these JE can interact with itself or other JE; and (ii) only one JE in AU and a single unbound JE in NSI are complementary, so that AU becomes non-covalently linked to NSI to produce a new NSI for use in subsequent cycles; (b) removing unbound AU; and (c) cyclic repetition of (a) and (b) to form a nanostructure. The new feature is that the complementary JE in at least one cycle are PNAs. Also described are nanostructures formed from many AU, comprising different nanostructures with a very wide range of potential applications, e.g. structural reinforcements (for aerogels, paper, plastics or cement, particularly as long fibres to improve tensile strength); identification (anti-counterfeiting) markers; enzyme or catalyst supports; assembly scaffolds; for construction of nanowires or nanocircuits; size markers for electron microscopy; molecular sieves and filters; substrates for optical and other surface coatings; scaffolds for solubilising enzymes or for trapping, protecting and delivering specific molecules; in high-density computer memories; as artificial zeolite for absorbing ions from water and for construction of new materials, including use in biosensors. PNAs are more homogeneous than inorganic nanoparticles with predictable geometry and stoichiometry. The present sequence represents a protein containing coiled coil dimerisation sequences that can be used for structural elements of assembly units, given in the exemplification of the present invention.  
 Sequence 961 AA;  
 Query Match 7.8%; Score 118; DB 7; Length 961;  
 Best Local Similarity 22.9%; Pred. No. 0.79;  
 Matches 72; Conservative 54; Mismatches 138; Indels 50; Gaps 12;  
 QY 7 DKTVEVVKNAIETADGALDLY-NKYLDQVTPWQTFDETIKELSRFKOYSQAASVVLGDI 65  
 Db 635 DKKEEVKNTLEQHDSIVTHYKNMIREQDLQLEELKQIISTLKONEQLOTAVTQVQSQI 694  
 QY 66 K-----TLLMDSQDK--YPEATQTVYVCGVATQLLAAYILLFDEYNEKAS-- 110  
 Db 695 QQHKQOYNLLKVQLGKDSQHQGPYTDGAQ---MNGVQPEEISR---LREEIEELKSNRE 747  
 QY 111 -----AQKDILIKVLDDGITYK--LNEAKSLVSSQSFNNASGKLALDLSQTLNDFSEK 162  
 Db 748 LLQSLAELKXDSLIENTLKSSQLSPGTNEOSSATAGDSEQIAELKQELATIKSQL-NSQSV 806



CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life cycle  
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site

XX Sequence 1211 AA;  
 CC Query Match 7.8%; Score 118; DB 5; Length 1211;  
 CC Best Local Similarity 19.3%; Pred. No. 1.1;  
 CC Matches 69; Conservative 53; Mismatches 112; Indels 124; Gaps 13;  
 QY 5 VADKTVVVK-----NAIETADGALDLYNKYLDOVIPWQTFDTIKELGRFKQVSOAA 58  
 Db 183 IIEESAGVLKVKRKAESIQKLDHTEDNLNRVEDLIDLEGRVPLKEEAAIAKEYKQLS 242  
 QY 59 S-----VLVDIKTLMDSD-----KYFEATQTVYVCGVATQLLAAYILLFDEYN 105  
 Db 243 KMEQSDVITVDDIDHTEDNQRLDERLNLKSOQAEK--GQQAQINQ-----LQKXK 296  
 QY 106 EKKASAKDILIKVLDDGDKITKNEAKSLLVSSQSFNNASGKLALDLSOLTND-----F 159  
 Db 297 GKRQNDYD-----IEKLN-----YELVKATENYQSLGNVLEERKKNQSETNARY 344  
 QY 160 SEKSSYFQSQVDKIRKEAVAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKN----- 213  
 Db 345 BEELNDLESQDISKNEK-----AKNEKLLADLKNKQKOLN 380  
 QY 214 -----KLKSVQNFPTLNTVQANKDI-----DAKLLKT 244  
 Db 381 KEVQELSLIYSDEQHEKLEIKNSYYTLMSQSVVNDIRFLEHTINEAKSRLD 440  
 QY 245 TEATAAIGEIKTETTRFYVDYDMLSLLEAKKWMINTCNEYQKRHGKTKLFEVPE 302  
 Db 441 SRLVE-----AFNQLKDIQQNITQTOKEYQS--SKKSMKEVQ 476

RESULT 14  
 AAR43563  
 ID AAR43563 standard; protein; 476 AA.  
 XX AC AAR43563;  
 XX 25-MAR-2003 (revised)  
 DT 09-JAN-2003 (revised)  
 DT 05-APR-1994 (first entry)  
 XX DE Hyaluronan receptor.  
 XX Hyaluronan binding protein; HA, RHAMM; mediated motility; wound; healing;  
 KW diagnosis; treatment; cell locomotion; tumour invasion; birth defects;  
 KW inflammatory disorder; Alzheimer's disease; dementia;  
 KW Parkinson's diseases; Huntington's disease; AIDS; diabetes; auto;  
 KW immune diseases; corneal dysplasia; hypertrophy; surgery; burns; strokes;  
 KW multiple sclerosis; depression; schizophrenia; CNJ; contraception;  
 KW in vitro fertilisation; embryo development.

XX Rattus sp.  
 XX WO9321312-A1.  
 XX 28-OCT-1993.  
 XX 13-APR-1993; 93WO-CA000158.  
 XX 09-APR-1992; 92GB-00007949.  
 XX (UYMA-) UNIV MANITOBA.  
 XX (MANI-) MANITOBA CANCER TREATMENT & RES FOUND.

XX Turley EA;  
 XX WPI; 1993-351722/44.  
 DR N-PSDB; AAO51212.  
 XX DNA encoding hyaluronan receptor - used to produce proteins and  
 PT antibodies for alteration of cell locomotion.  
 XX Claim 7; Fig 23; 88pp; English.  
 XX The sequence is that encoded by a cDNA clone encoding the hyaluronan  
 CC receptor (HARC). The sequence was obt'd. by screening a 3T3 library in  
 CC lambda gt11 with antibodies to HARC. A clone of 1.9 kb was obtained and  
 CC used to rescreen the library to obtain the full length, 2.9 kb clone. HA  
 CC is down regulated in stationary normal cells and is only expressed in  
 CC situations where cell motility is desired, e.g. in wound healing, in  
 CC response to growth factors and in chemotaxis by white blood cells. HA may  
 CC be used for diagnosis and treatment of diseases involving cell  
 CC locomotion, e.g. tumour invasion, birth defects, acute and chronic  
 CC inflammatory disorders, Alzheimer's and other forms of dementia, AIDS,  
 CC diabetes, autoimmune diseases, corneal dysplasias and hypertrophies,  
 CC burns, surgical incisions and adhesions, strokes, multiple sclerosis,  
 CC depression/schizophrenia related to neuronal growth and pain states  
 CC involving nerve sprouting; also in CNJ and spinal cord regeneration,  
 CC contraception, in vitro fertilisation and embryo development. See also  
 CC AAR46548-51. (Updated on 09-JAN-2003 to add missing OS field.) (Updated  
 CC on 25-MAR-2003 to correct PN field.)

XX Sequence 476 AA;  
 CC Query Match 7.8%; Score 117.5; DB 2; Length 476;  
 CC Best Local Similarity 20.8%; Pred. No. 0.33;  
 CC Matches 69; Conservative 73; Mismatches 125; Indels 65; Gaps 18;  
 QY 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDOVIPWQTFDTIKELSRFKQVSOASV 60  
 Db 99 LDNLREKEVELEKHTARQAAILIAQEKYIDTA---QSLRVVTAQLESVQEKYNDTAOS 155  
 QY 61 LVGDIKTLMDSDQKYEATQTVYVCGVATQLLAAYILLFDEYNEKKAQAQDILIKVL 120  
 Db 156 L-RDVTQALESBOEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 205  
 QY 121 DDGITKLENAQKSL--VSSQSFNNASGKLAL--DSQLTN-DPSEKSYFQSQVDKIRKEA 177  
 Db 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLLENLTLOEKVAMAEKSVEDVQOQI 265  
 QY 178 YAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNF-----TTLS 226  
 Db 266 LTAESTNQ-----EYA-----RMVDLQNRSTLKEEIKETSSFLEKITDLK 308  
 QY 227 NTVVKQANKDI-----DAAKLKLTTEI-----AAIGEIKTETTRFYVD-YDD 268  
 Db 309 NQLRQDDEDFRQLEEKGRKTAENVMVTELTMEINKWALLYELYEKTPFQQLDAFEA 368  
 QY 269 LMSLLKE--AAKWMINTCNE-YQKRHGKTKL 297  
 Db 369 EQQALLNEHGATQEQQLNKIRDSYAQLLGHQNL 400

RESULT 15  
 ABB49720  
 ID ABB49720 standard; protein; 927 AA.  
 XX AC ABB49720;  
 XX 05-FEB-2002 (first entry)  
 XX Listeria monocytogenes protein #2424.  
 XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.

OS Listeria monocytogenes.  
PN WO200177335-A2.  
XX 18-OCT-2001.  
XX 11-APR-2001; 2001WO-FR001118.  
XX 11-APR-2000; 2000FR-00004629.  
XX (INSP ) INST PASTEUR.  
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
PI Dusserget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Vose H;  
XX WPI; 2002-010914/01.  
DR Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and related  
PT polypeptides.  
XX Claim 6; SEQ ID NO 2425; 192pp; French.  
XX The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 927 AA;  
Query Match 7.8%; Score 117.5; DB 5; Length 927;  
Best Local Similarity 21.1%; Pred. No. 0.83;  
Matches 73; Conservative 33; Mismatches 141; Indels 99; Gaps 11;  
QY 3 EIVADKTVVVKNAETADGALDLYNKVLDOVIPWOTFDETI-----KELSRFKQVYQAA 58  
DB 166 EATPDKIKESGDGPAADGS-----GKIDGLVKVSKQGNKTIISTNLKTLADSSLTFPKDGA 221  
QY 59 SVL-----VGDIK-----TLLMDSQDKYF 77  
DB 222 NTLVGLKTYTDGNTAAAGDKLKNAGVSTLAAGVGLKDGVAALDGGATKLASGVSTYT 281  
QY 78 EATQVYEWCVATQALLAAVILLFDEYNEKKAQKDILIKVLDGDTIKLNEAQSLLVS 137  
DB 282 SGVDTL--AGGINQAVTGSTALSDGLNKNMS-----VPTLASGITQLNNGQKSLATG 332  
QY 138 SQSFNNASGKLLA-----LDQLTN-----DFSEKSSYFQSDVKIRKAYAGAAQVWA 187  
DB 333 LDSLVDGSKNLKSLAGLXELDGNLTDKQCKIAQLKQGMNDLQOQIDQLNQSVNGEDAA----- 388  
QY 188 GPFGLIISYSIAAGVWEGKILPELKNKLSQNFVFTLSNTVKQANKDIDAALKLITTEI 247  
DB 389 -----LAKQLATLQKSLDSLQNLGLTTFIKS-----NANFDAEAIKSKINATA 429

QY 248 AAI GEIKTETTRFYVDYDMLSLKEAAKQKQKNTCNEYQKRHG 293  
DB 430 GVS AEDKQKI-----IDAIQADLDKQKATQVATVQEQSQG 467

RESULT 16  
AAG82283  
ID AAG82283 standard; protein; 885 AA.  
XX AAG82283;  
XX 03-SEP-2001 (first entry)  
XX S. epidermidis open reading frame protein sequence SEQ ID NO:1660.  
XX Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;  
XX endocarditis.  
XX Staphylococcus epidermidis.  
XX WO200134809-A2.  
XX 17-MAY-2001.  
XX 09-NOV-2000; 2000WO-US030782.  
XX 09-NOV-1999; 99US-0164258P.  
XX (GLAX ) GLAXO GROUP LTD.  
XX Kimmerly WJ;  
XX WPI; 2001-316495/33.  
XX N-PSDB; AAH53133.  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis.  
XX Claim 18; Page 462; 2188pp; English.  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG8120, from Staphylococcus epidermidis. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH5091 to AAH5098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX Sequence 885 AA;  
Query Match 7.7%; Score 116; DB 4; Length 885;  
Best Local Similarity 20.0%; Pred. No. 1;  
Matches 63; Conservative 42; Mismatches 92; Indels 118; Gaps 12;  
QY 42 ETIKELSRFKQVYQAA-----VLVGDITKLLMDSQ-----KYFRATQVYEWCG 88  
DB 2 EPLKEEAIAKEVYKQLSKEMEQQSDVITVSDIDHYTEDNQRLDLRLNHLKSSQAQKE--G 59  
QY 89 VATQLLAAYILLFDEYNEKKAQKDILIKVLDGDTIKLNEAQSLLVSQSFNASGKL 148

|    |                       |  |          |            |                  |            |          |     |
|----|-----------------------|--|----------|------------|------------------|------------|----------|-----|
|    | Best Local Similarity | 23.3%; Pred.   | No. 1.4; | Mismatches | 67; Conservative | 55; Indels | 46; Gaps | 13; |
| Qy | 3                     | EIVADKTVVVGNAIETADGALDLYNKYLDVPWQTDETIKELSRPKQYSQAASVLV        | 62       |            |                  |            |          |     |
| Dd | 434                   | KLIADWLLEINQTSINKAGSLDRIGIDKA-----RNASPQAEIRGLGBKEADS--W       | 485      | :          | :                | :          | :        | :   |
| Qy | 63                    | GDIKTLLMDSODKYFEATQTVYWCVAQL--LAAYILLFDEYNKKAKAQKDILIKVL       | 120      | :          | :                | :          | :        | :   |
| Dd | 486                   | SGKSLLMDLINDKY---QKIINEKSAVEVELNSLNKLNTQTRRQATLQRK--LORI       | 540      | :          | :                | :          | :        | :   |
| Qy | 121                   | DDGITKLINEAQS--LAVSSOSFNNASGKLIALDSLDTNDFSEKSYSPSQVDVKIRKEAY   | 178      | :          | :                | :          | :        | :   |
| Dd | 541                   | NDAENKLPALORSVDIIIDNFPRKFNEHRLTRTANSLTHRDRIQRSALNRMTPLEREELD   | 600      | :          | :                | :          | :        | :   |
| Qy | 179                   | AGA----AAGVAGPFGIHISYSTAAGVGEK-----LPE--LKKNKLSVNFFFTTL        | 225      | :          | :                | :          | :        | :   |
| Dd | 601                   | AAADDIVNKIIGAPGI VPSSELIPDLGVKRAGFKTORTLINIPDERIKOYLSDSVNI--VM | 658      | :          | :                | :          | :        | :   |
| Qy | 226                   | SNTVTQANKDDAAKLLTYEIAAI---GRICKTETTTFPYDYDDL M                 | 270      | :          | :                | :          | :        | :   |
| Dd | 659                   | ENYIQV-----APSIETAFKFRDMNQIKAITE-----EYNOLI                    | 695      | :          | :                | :          | :        | :   |

|                       |  |
|-----------------------|--|
| XX                    | ADL05067;  |
| XX                    | AC   |
| XX                    | ADL05067;  |
| XX                    | 06-MAY-2004 (first entry)  |
| DT                    | XX   |
| DE                    | M. catarrhalis protein #833.   |
| XX                    | Moraxella catarrhalis; infection.  |
| KW                    | Moraxella catarrhalis;   |
| XX                    | OS   |
| XX                    | Moraxella catarrhalis.   |
| XX                    | US6673910-B1.  |
| PN                    | XX   |
| PD                    | 06-JAN-2004.   |
| XX                    | XX   |
| PP                    | 04-APR-2000; 2000US-00540236.  |
| XX                    | XX   |
| PR                    | 08-APR-1999; 99US-0128416P.  |
| XX                    | XX   |
| PA                    | (GENO-) GENOME THERAPEUTICS CORP.  |
| XX                    | XX   |
| PI                    | Breton GL;   |
| XX                    | XX   |
| DR                    | WPI: 2004-178127/17.   |
| DR                    | N-FSDB; ADL03147.  |
| XX                    | XX   |
| PT                    | New nucleic acid encoding a Moraxella catarrhalis polypeptide, useful for preparing a composition for diagnosing, preventing or treating infection caused by Moraxella catarrhalis.  |
| PS                    | Disclosure; SEQ ID NO 2753; 429pp; English.  |
| XX                    | XX   |
| CC                    | The invention relates to an isolated nucleic acid encoding an Moraxella catarrhalis polypeptide. The nucleic acid is useful for preparing a composition for diagnosing, preventing or treating infection caused by Moraxella catarrhalis. The present sequence represents the amino acid sequence of a M. catarrhalis protein. |
| XX                    | XX   |
| SQ                    | Sequence 718 AA;   |
| Query Match           | 7.7%; Score 115.5; DB 8; Length 718;   |
| Best Local Similarity | 21.0%; Pred. No. 0.86;   |
| Matches               | 70; Conservative 57; Mismatches 130; Indels 77; Gaps 1   |
| YQ                    | 6 AKTVEVVKNAETAGDALDLYNKVLDOVIPWQTDFDTIKELSRFKQE-----YSQA 57   |

[illegible]

Db 103 ATEKLEVAKEATQ-----DKVEKTSQSLVEDIKDKAQSLOEDAADTVEALKQA 149  
Qy 58 ASVLVGDIKTLMLMSQDKYFEATQTVYEWG-----GVATQLLAAYILLFDEYN--- 105  
Db 150 AS---DKVETTKABAQSLKDDATQT-PESAQAQVEGKVEAIKEQVLQVDSLKDDTTQDN 205  
Qy 106 ---EKKASAKDILILIKVLDDGITKLNQAQSLVSSQSP-NNASGKL-----LALD--SQ 154  
Db 206 TDQOEKQTLKDKAVQAATAAKRKVEDVDVVKHTTESFRNTAGSKIDEIKQAADVKTTEE 265  
Qy 155 LTNDSEKSSYFQSDVKIRKEAYAGAAAGVVAGPFGLIISYSIAAGVVEGKLIPELKNK 214  
Db 266 VKSLSQKADALKSSGSELKQTAT-AANDAITEAQAADVSGSVAADSAQSTAQSAKDX 324  
Qy 215 LKSVQNFPTLSNTVQKQNDIDAQKLTTEIAAIGIKETETETTRFYVDYDMLSL 274  
Db 325 L-----NQLFEQKGALDEKQVQELGKFKGATEKINAVSENVDLATQVI 367  
Qy 275 KEAKKMINTC-----NEVQKRGHKKTL 297  
Db 368 KEEAQAQLQTNQBSLQAQKAGEBYDATHDKGL 401

## RESULT 19

ABBS9344

ID ABB59344 standard; protein; 2056 AA.

XX AC ABB59344;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 4824.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN W0200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL03447.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX PT genes from Drosophila and for elucidating cell signaling and cell-cell

XX PT interactions.

XX PS Disclosure; SEQ ID NO 4824; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent

XX CC capable of detecting 1000 or more genes from Drosophila. The invention is

XX CC useful in developmental biology and in elucidating cell signalling and

XX CC cell-cell interactions in higher eukaryotes for the development of

XX CC insecticides, therapeutics and pharmaceutical drugs. The invention

XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-

XX CC ABB572072). The sequence data for this patent did not form part of the

XX CC printed specification, but was obtained in electronic format directly

XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX CC Sequence 2056 AA;

Query Match 7.68; Score 115; DB 4; Length 2056;  
Best Local Similarity 21.3%; Pred. No. 4;  
Matches 80; Conservative 48; Mismatches 107; Indels 140; Gaps 16;  
Qy 14 KNALETADGALDLYNKYL-----DQVIPWOTFDETIKELSRFK 51  
Db 1302 KTVLEKAKGTLEAENADLATELSVNSSRQENDRRKQAESQIAELQV---KLAIEERAR 1358  
Qy 52 QEYSQAASVLVGDITKTLMLMSQDKYFEATQTVYEWGCVATQLLAAYILLFDEYNK----- 107  
Db 1359 SELQEKCTKLQEAENITNQLAEALKASAAVKSASNMESQLEAQQLLBEETRQKLGSL 1418  
Qy 108 -----KASAQKDI-LIKVLDDGIT 125  
Db 1419 SKLRQISEKALQEQLEEDDEAKRNYERKLAEVTTQMQEIKKKAEBDADLAKELEGKK 1478  
Qy 126 KLN-----BAQ-KSLLVSGSQFNNSGKL-----LALDSQLTNDFS-EKSYFQSQ 169  
Db 1479 RLKNDIEALERQVKELIAQNDRDLDSKKKTKQSELEDATIEAORTKVLELEKK---QKN 1535  
Qy 170 VDKIRKEAYAGAAAGVVAGPFGLIISYSIA-----AGVVEGKL----- 207  
Db 1536 FDKILAEEKA-----ISQIAQERDTABREAREKETKVLVSRELDEAFDK 1581  
Qy 208 IPELKNKLSVQNFPTLSNTVQKQNDIDAQKLTTEIAAIGIKETETETTRFYV 264  
Db 1582 IEDLENKRTKLQNELDDLANTQGTADKNVHELEKAKKRALESQLA---ELKAQNEELE--- 1635  
Qy 265 DYDDLMLSLLEAKA 279  
Db 1636 --DDLQLT---EDAK 1645

## RESULT 20

ABM68103

ID ABM68103 standard; protein; 582 AA.

XX AC ABM68103;

XX DT 20-NOV-2003 (first entry)

XX DE Photorhabdus luminescens protein sequence #1200.

XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;

XX KW detection; food; gene expression; plant; animal; microorganism; toxin;

XX KW antibiotic; biopesticide; virulence factor; disease model; plague;

XX KW whooping cough.

XX OS Photorhabdus luminescens.

XX PN W0200294867-A2.

XX PD 28-NOV-2002.

XX PF 07-FEB-2002; 2002WO-IB003040.

XX PR 07-FEB-2001; 2001FR-00001659.

XX PA (INSP ) INST PASTEUR.

XX PA (CNRS ) CNRS CENT NAT RECH SCI.

XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;

XX PI Buchrieser C;

XX DR WPI; 2003-148459/14.

XX PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,

XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX PS Claim 2; SEQ ID NO 1200; 1205pp; French.

XX CC The invention relates to the isolation of genes and their encoded

XX CC proteins from Photorhabdus luminescens. The isolated sequences are



sources of probes and primers for detecting the genome of *P. luminescens* and related species; to study polymorphisms, for gene analysis and for detection/amplification of the genes. Antibodies (Ab) raised against the polypeptides encoded by the genes are used for detection/identification of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that carry a gene-containing vector are used to select compounds that modulate, regulate, induce or inhibit expression of the genes in plants, animals or microorganisms other than *P. luminescens* and are able to alter response or sensitivity to toxins and antibiotics produced by *P. luminescens*. Cells transformed to express the genes are useful for recombinant production of the proteins, particularly toxins and antibiotics useful as insecticides, bactericides and fungicides. The genes, proteins, vectors containing the genes and Ab are also useful therapeutically to treat microbial infection by bacteria or fungi that are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as biopesticides. Other uses of the genes and the proteins are as virulence factors and for identifying targets of human diseases for which *P. luminescens* is a model (particularly plague and whooping cough). This sequence represents one of the isolated *P. luminescens* proteins

Sequence 582 AA:

|                       |       |  |                 |             |
|-----------------------|-------|--|-----------------|-------------|
| Query Match           | 7.6%  | Score 114.5;   | DB 6;           | Length 582; |
| Best Local Similarity | 20.2% | Pred. NO. 0.79;  |                 |             |
| Matches               | 64;   | Conservative 58;   | Mismatches 118; | Indels 77;  |
|                       |       |  |                 | Gaps 11     |
| QY                    | 17    | LETADGALDLYNKYLDQVIPWQTFDETIKELSRFQEQYSQAASVLVIGDIKTLMDSDQKY | 76              |             |
| DB                    | 133   | LETPTPAYFYFPFYIDQIKSHSSPWNSENLOQSNWKGP LIKYFTGYLKPHEHFIENI   | 192             |             |
| QY                    | 77    | FEATQTVYEWCGVATQLAAAYILLFDEYNEKKAQAQOILIKV                   | 126             |             |
| DB                    | 193   | YEYSIEIKESAHIKEKFSQAVEVIVDNTVDSPTILDNNDFIQIENIKNELYDLIDYQTR  | 252             |             |
| QY                    | 127   | LNEAQSKLLVSSQSFNNASGKLLALDS--OLTNDP-----SEKSSVFQ-----        | 167             |             |
| DB                    | 253   | LYDAQAT--ITSNIYDLESQYALATISANELEADYKFAVESIPTDYLECPCLGTLHDNSL | 310             |             |
| QY                    | 168   | -----SQVDKIRKEAYAGAAGVVAGPFGLLIYSIAAGVVEGKLIIDELKNKLKSVQN    | 220             |             |
| DB                    | 311   | PNRAILLSEKDSLLNEA-----NSIASKIV-----ELKSSLSNLSNE              | 346             |             |
| QY                    | 221   | FFTTLSTNTVQANK-----DIDAAKLKTWETIAAIGIKTE--TETTRFVVDVDDLMLSL- | 273             |             |
| DB                    | 347   | DAQFITNEIERINNKYITDDDAVKGLIAQV--IDTLSTENVSKNITQIKIDNEDLNISKA | 404             |             |
| QY                    | 274   | -----LKEAAKKOMINT  | 284             |             |
| DB                    | 405   | NSSIKELKKDQKKLLST  | 421             |             |

|           |   |
|-----------|---|
| RESULT 21 |   |
| ABU42396  |   |
| ID        | ABU42996 standard; protein; 1189 AA.                                    |
| XX        |   |
| AC        | ABU42996;   |
| XX        |   |
| DT        | 19-JUN-2003 (first entry)   |
| XX        |   |
| DE        | Protein encoded by Prokaryotic essential gene #28523.                   |
| XX        |   |
| KW        | Antisense; prokaryotic essential gene; cell proliferation; drug design. |
| XX        |   |
| OS        | Staphylococcus epidermidis.   |
| XX        |   |
| PN        | WO200271183-A2.   |
| XX        |   |
| PD        | 03-OCT-2002.  |
| XX        |   |
| PF        | 21-MAR-2002; 2002WO-US009107.   |
| XX        |   |
| PR        | 21-MAR-2001; 2001US-00815242.   |
| PR        | 06-SEP-2001; 2001US-00948993.   |
| XX        |   |

25-OCT-2001; 2001US-0342923P.  
08-FEB-2002; 2002US-00072851.  
06-MAR-2002; 2002US-0362659P.  
(ELIT-) ELITRA PHARM INC.  
Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen Ku, Zyskind JW;  
Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
WPI; 2003-029926/02.  
N-PSDB; ACA46866.  
New antisense nucleic acids, useful for identifying proteins or screening  
for homologous nucleic acids required for cellular proliferation to  
isolate candidate molecules for rational drug discovery programs.  
Claim 25; SEQ ID NO 70920; 1766pp; English.  
The invention relates to an isolated nucleic acid comprising any one of  
the 6213 antisense sequences given in the specification where expression  
of the nucleic acid inhibits proliferation of a cell. Also included are:  
(1) a vector comprising a promoter operably linked to the nucleic acid  
encoding a polypeptide whose expression is inhibited by the antisense  
nucleic acid; (2) a host cell containing the vector; (3) an isolated  
polypeptide or its fragment whose expression is inhibited by the  
antisense nucleic acid; (4) an antibody capable of specifically binding  
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
proliferation or the activity of a gene in an operon required for  
proliferation; (7) identifying a compound that influences the activity of  
the gene product or that has an activity against a biological pathway  
required for proliferation, or that inhibits cellular proliferation; (8)  
identifying a gene required for cellular proliferation or the biological  
pathway in which a proliferation-required gene or its gene product lies  
or a gene on which the test compound that inhibits proliferation of an  
organism acts; (9) manufacturing an antibiotic; (10) profiling a  
compound's activity; (11) a culture comprising strains in which the gene  
product is overexpressed or underexpressed; (12) determining the extent  
to which each of the strains is present in a culture or collection of  
strains; or (13) identifying the target of a compound that inhibits the  
proliferation of an organism. The antisense nucleic acids are useful for  
identifying proteins or screening for homologous nucleic acids required  
for cellular proliferation to isolate candidate molecules for rational  
drug discovery programs, or for screening homologous nucleic acids  
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
*K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
the target prokaryotic essential genes. Note: The sequence data for this  
patent did not form part of the printed specification, but was obtained  
in electronic format directly from WIPO at  
frn.wipo.int/pub/published/pct sequences

|                       |        |   |   |                |
|-----------------------|--------|---|---|----------------|
| Query Match           | 7.6%;  | Score 114.5;  | DB 6;   | Length 1189;   |
| Best Local Similarity | 20.1%; | Pred. No. 2.1;  |   |                |
| Matches               | 67;    | Conservative  | 45;   | Mismatches 96; |
|                       |        |   |   | Indels 125;    |
|                       |        |   |   | Gaps 13        |
| QY                    | 5      | VADKTVEVK-----  | NAIETADGALDLYNKYLQDVIPWQTFDETIKLSRFKQBYSOAA       | 58             |
|                       |        | : : : : :   | : : : : :   | : : : : :      |
| Db                    | 161    | IIIESAGVLVKYKRKAESIQKLHTELDNLNRVEDILYDLGRVEPLKEAAIAKAYKOLS    |   | 220            |
|                       |        | : : : : :   | : : : : :   | : : : : :      |
| QY                    | 59     | S-----  | VLVGDIKTLMLDSQKYPEATQTVYEWG-----                  | GVATQLLAA 96   |
|                       |        | : : : : :   | : : : : :   | : : : : :      |
| Db                    | 221    | KEMEQSDVIVTVSDI-----  | DHYTEDNQRLDERLNHLKSQQAQKESQQAQINQLLQR             | 272            |
|                       |        | : : : : :   | : : : : :   | : : : : :      |
| QY                    | 97     | YILLFDEYNEKKASQAQKDILIKVLDDGTIKLNEAQKSLIVSSQSFNNAAGKLLALDSQLT |   | 156            |
|                       |        | : : : : :   | : : : : :   | : : : : :      |
| Db                    | 273    | Y-----  | KGKQQN-----                                       | DYDIEKLN-----  |
|                       |        | : : : : :   | : : : : :   | : : : : :      |
| QY                    | 157    | ND-----   | PSEKSYFQSQVDKIRKEAYAGAAGVVGAFGLIISYSIAAGVVEGKLIPE | 210            |
|                       |        | : : : : :   | : : : : :   | : : : : :      |
| Db                    | 314    | NQSETNARYEBELDNLESQIDISIKNEK-----                             |   | AQNEKLLAD 349  |
|                       |        | : : : : :   | : : : : :   | : : : : :      |
| QY                    | 211    | LKN-----  | KLKSVQNQFTTLLSNVTKQANKDI-----                     | 236            |





Query Match 7.6%; Score 114; DB 5; Length 796;  
Best Local Similarity 24.9%; Pred. No. 1.3;  
Matches 68; Conservative 44; Mismatches 119; Indels 42; Gaps 10;

QY 42 ETIKELSRPKQYEQYQASVLDGDKITLLMSQDKYFEATQVYEWCGVATQLLAAVTLF 101  
DB 471 EQILKAKETSTAEFFSSVLNEMIEKLEKSEIDDEYTAIAV-----GLESLTA---MR 522

QY 102 DEYNEKASAKOILIKVLDGDKITLLMSQDKYFEATQVYEWCGVATQLLAAVTLF 153  
DB 523 EEPF--KASSEHLMHPVLEKIEKLEEFNTRLTDPNYESLSKSLNMLRDFSRKAAS 580

QY 154 QLTNDFSEKSYFQSQVDK-----IRKEAYAGAAVGVVAGPFGILISYSIAAGVVE 204  
DB 581 EATSLKKEINKRFQCAVDREIREKVEAKAEVASSGASSFDELPDALKKVLTKGEVE 640

QY 205 GKILPELKN--KLKSVQNFFTLSTNTVKQANKDIDRAKLTTEIAA-IGEIKTETETT 260  
DB 641 AEMAGVLKSMGLELDVAVKQKDTAEQIYAANENLOEKLEKLNQEIITSKIEEVVTRPEIK 700

QY 261 RFYVDYDMLSLK-BAK--XMTNCTNEYOK 290  
DB 701 S-----MVELLKVETAKASKTPGVTEAYOK 725

RESULT 24  
ADC94317  
ID ADC94317 standard; protein; 1196 AA.

XX ADC94317;  
XX  
XX 01-JAN-2004 (first entry)  
XX  
XX E. faecium protein sequence SEQ ID 3944.

XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
XX abdominal-pelvic infection.  
XX Enterococcus faecium.  
XX  
XX US6583275-B1.  
XX  
XX 24-JUN-2003.  
XX  
XX 30-JUN-1998; 98US-00107532.  
XX  
XX 02-JUL-1997; 97US-0051571P.  
XX  
XX 14-MAY-1998; 98US-0085598P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2003-799836/75.  
XX  
XX N-PSDB; ADC90663.  
XX  
XX New isolated nucleic acid derived from Enterococcus faecium encoding an  
XX Enterococcus faecium polypeptide useful for detection, prevention and  
XX treatment of a pathological condition resulting from a bacterial  
XX infection.  
XX  
XX Example 1; SEQ ID NO 3944; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from  
XX Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
XX one of 10 fully defined sequences given in the (or comprising 40  
XX sequential nucleotides chosen from any of the nucleic acids, its  
XX complement or sequences hybridising to it). Also included are a  
XX recombinant vector comprising the nucleic acid operably linked to  
XX transcription regulatory element, a cell comprising the vector and a  
XX single-stranded probe comprising the nucleic acid. The nucleic acids are  
XX chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.

The nucleic acids is useful for diagnosing pathological conditions  
resulting from E. faecium bacterial infection (e.g. urinary tract  
infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
infection) and for screening drugs such as agonists and antagonists. The  
nucleic acid is useful for recombinant production of Candida albicans -  
derived peptides or antisense polypeptides. Pharmaceutical compositions  
and vaccines containing the nucleic acid are useful for preventing or  
treating Enterococcus faecium infections. The present sequence represents  
one of the disclosed E. faecium proteins.

XX Sequence 1196 AA;  
SQ

Query Match 7.5%; Score 113.5; DB 7; Length 1196;  
Best Local Similarity 20.0%; Pred. No. 2.6; 111; Indels 107; Gaps 12;  
Matches 67; Conservative 50; Mismatches 50;

QY 5 VADKTVEVVK-----NAIETADG---ALDLYNKYLDQVIPWQTFDETIKELSRPK 51  
DB 163 IFETAGVLKYQKQKKAQKLFETEDNLSRVODIHELEQUTPLAAQSEAAKEFLRLK 222

QY 52 QEYSQA-ASVYVDGDKITLLMSQDKYFEATQVYEWCGVATQLLAAVTLFDEYNEKKA 110  
DB 223 ETLTQTDVSLMVAEIKTKKD-----WDNKQAOL----- 251

QY 111 AQKDILIKVLDGDKITLLMSQDKYFEATQVYEWCGVATQLLAAVTLFDEYNEKKA 170  
DB 252 AKPNELGLSESI-----QEESILAKORKEAQAADRLIEKQVLLDSEKLTQTEGQK 307

QY 171 DKIRKEAYAGAAVGVVAGPFGILISYSIAAGVVEGKLIPELKNKLSVQNFFTLSTNTVK 230  
DB 308 DVLQERT-----KHTQKSQSEYQTSLAQAOK 333

QY 231 QA-----NKDIDAAKLTTEI--AIGIKTETE-----TTRFYVDYDMLSLK 275  
DB 334 KVHFPEKLOESLMKAAAEKETEIQKABANLIKQOELEKYQKSTKELLAELRDQYVDLMQ 393

QY 276 EAAKMMINTCNEYQKRH-----GKKTLEFVEDEV 303  
DB 394 EQA--AVGNELKYLERYIQETAKSKQTLAKQSEV 426

RESULT 25  
ABG80428  
ID ABG80428 standard; protein; 458 AA.

XX ABG80428;  
XX  
XX 29-NOV-2002 (first entry)  
XX  
XX Moraxella catarrhalis Tola protein.  
XX  
XX Hyperblebbing; Gram-negative bacterium; genetically modified; Tol gene;  
XX peptidoglycan-binding; peptidoglycan-associated site; outer-membrane;  
XX bacterial infection; vesicle-shedding; Bleb; filter sterilised;  
XX detergent; deoxycholate; homogeneity; antibacterial; vaccine; Tola; Tola;  
XX OmpCD; xOmpA; PAL-1; PAL-2.  
XX  
XX Moraxella catarrhalis.  
XX  
XX WO200262378-A2.  
XX  
XX 15-AUG-2002.  
XX  
XX 08-FEB-2002; 2002WO-EP001361.  
XX  
XX 08-FEB-2001; 2001GB-00003171.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
XX Berthet FJ, Denoel P, Neyt CA, Poolman J, Thonnard J;  
XX WPI; 2002-657509/70.  
XX  
XX N-PSDB; ABS66198.













CC antiproliferative or antiinflammatory compound. The methyl donor compound  
CC acts synergistically with the antiviral, antiproliferative or  
CC antiinflammatory compound to provide an enhanced therapeutic effect with  
CC minimal side effects. In the antiinflammatory compositions of the  
CC invention, the amount and the frequency of administration is lower than  
CC that required for the antiinflammatory compound in the absence of the  
CC methyl donor compound. Compositions of the invention can be used in the  
CC treatment of viral disease, proliferative diseases (e.g., cancer), and  
CC inflammatory diseases such as multiple sclerosis. The present sequence  
CC represents a peptide fragment of murine RHAMM (receptor for hyaluronic  
CC acid mediated motility) which is specifically claimed as the  
CC antiinflammatory compound in an antiinflammatory composition of the  
CC invention for the treatment of multiple sclerosis  
XX  
SQ Sequence 435 AA;  
Query Match 7.4%; Score 111; DB 6; Length 435;  
Best Local Similarity 20.5%; Pred. No. 1.1;  
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;  
QY 1 MTEIVADKTVVKNVNAITAGALDLYNKYLDQVTPWQTFDETIKELSRFKQESQASV 60  
Db 99 LDNLLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESQEKYNDTAQS 155  
QY 61 LVGDIKTLMDSQDKYFEATQTVYVWCGVATQLLAAYILLDFEYNEKASAKQDILIKVL 120  
Db 156 L-RDVTAQLESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQSRLDVTQA-L 205  
QY 121 DDGITKLNEAQKSL-L-VSSQSFNNASGKLAL-DSQLTN-DPSEKSSYFQSDVKIRKEA 177  
Db 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLENLTLOEKVMAEKSVEDVQQOI 265  
QY 178 YAGAAAGVAGPFGLLIISYIAAGVVEGKLIPELKNK-----LKSQVNF-----TTLS 226  
Db 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEEIKETSSFLEKITDLK 308  
QY 227 NTVQANKDI-----DAKLLKLTTEIAAIGEIKETETTRFFVVDYDDL 269  
Db 309 NQLRQODEDFRQLEEKGRKTAENVMTELT-----MEINKWRLLYEEL 353  
RESULT 34  
ADC02471  
ID ADC02471 standard; protein; 435 AA.  
XX  
AC ADC02471;  
XX  
XX 18-DEC-2003 (first entry)  
XX  
XX Mouse RHAMM V3 protein.  
XX  
XX RHAMM; Receptor hyaluronic acid mediated motility;  
XX cell surface target protein; AP-1; activating protein-1;  
XX response-to-injury; vaccine; inflammatory neurological disorder;  
XX diabetes mellitus; arthritis; inflammatory dermatosis;  
XX inflammatory bowel disease; cancer; kidney fibrosis;  
XX inflammatory lung disease; obesity; lupus; cardiovascular disease; wound;  
XX multiple sclerosis; Parkinson's disease; Alzheimer's disease;  
XX alpha helical peptide; mouse.  
XX  
OS Mus sp.  
XX  
XX  
XX US2003100490-A1.  
XX  
XX 29-MAY-2003.  
XX  
XX 15-OCT-2001; 2001US-00978309.  
XX  
XX 01-APR-1999; 99US-0127457P.  
XX 03-APR-2000; 2000US-00541522.  
XX 05-OCT-2000; 2000US-00685010.  
XX  
XX (CRUZ/) CRUZ T.

(PAST/) PASTRAK A.  
(TUR/) TURLEY E A.  
XX Cruz T, Pastrak A, Turley EA;  
XX WPI; 2003-755209/71.  
XX New polypeptide for diagnosing, preventing or treating inflammatory  
XX neurological disorders, cancer, obesity, lupus or diabetes mellitus, S7  
XX comprises an amino acid sequence of P16, P32, or murine or human S3, S7  
XX or V2.  
XX Disclosure; SEQ ID NO 80; 110pp; English.  
XX  
XX The invention relates to a polypeptide comprising an amino acid sequence  
XX selected from the sequences of P16, P32, murine S3, human S3, murine S7,  
XX human S7, murine V2 and human V2, all derived from the protein RHAMM  
XX (receptor hyaluronic acid mediated motility) a cell surface target  
XX protein that is required for activation of the AP-1 (activating protein-  
XX 1) pathway on response to injury. Also included are an antibody that  
XX binds to the polypeptide cited above, a pharmaceutical or vaccine  
XX compositions for treating an inflammatory neurological disorder or  
XX for the antibody cited above) and treating an inflammatory neurological  
XX diabetes mellitus (comprising the above amino acid sequence or an antigen  
XX disorder, arthritis, inflammatory dermatosis, inflammatory bowel disease,  
XX cancer, kidney fibrosis, inflammatory lung disease, obesity, lupus,  
XX cardiovascular disease, diabetes mellitus or wounds (comprising  
XX administering to a patient the above polypeptide or antibody. In the  
XX above method, the inflammatory neurological disorder is multiple  
XX sclerosis, Parkinson's disease or Alzheimer's disease. The polypeptide or  
XX antibody may also be used for diagnosing or preventing the above  
XX diseases. The present sequence is a mouse RHAMM active protein fragment.  
XX  
SQ Sequence 435 AA;  
Query Match 7.4%; Score 111; DB 7; Length 435;  
Best Local Similarity 20.5%; Pred. No. 1.1;  
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;  
QY 1 MTEIVADKTVVKNVNAITAGALDLYNKYLDQVTPWQTFDETIKELSRFKQESQASV 60  
Db 99 LDNLLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESQEKYNDTAQS 155  
QY 61 LVGDIKTLMDSQDKYFEATQTVYVWCGVATQLLAAYILLDFEYNEKASAKQDILIKVL 120  
Db 156 L-RDVTAQLESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQSRLDVTQA-L 205  
QY 121 DDGITKLNEAQKSL-L-VSSQSFNNASGKLAL-DSQLTN-DPSEKSSYFQSDVKIRKEA 177  
Db 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLENLTLOEKVMAEKSVEDVQQOI 265  
QY 178 YAGAAAGVAGPFGLLIISYIAAGVVEGKLIPELKNK-----LKSQVNF-----TTLS 226  
Db 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEEIKETSSFLEKITDLK 308  
QY 227 NTVQANKDI-----DAKLLKLTTEIAAIGEIKETETTRFFVVDYDDL 269  
Db 309 NQLRQODEDFRQLEEKGRKTAENVMTELT-----MEINKWRLLYEEL 353  
RESULT 35  
ABR56999  
ID ABR56999 standard; protein; 477 AA.  
XX  
XX ABR56999;  
XX  
XX 05-AUG-2003 (first entry)  
XX  
XX Mouse RHAMM V-2 region amino acid sequence SEQ ID NO:77.  
XX  
XX RHAMM; receptor hyaluronic acid mediated motility; hyaluronic acid; HA;  
XX neuroprotective; vasotropic; antidiabetic; nootropic; antiarthritic;  
XX antiparkinsonian; antiinflammatory; osteopathic; dermatological; vaccine;



|   |   |
|---|---|
| KW  | antiarteriosclerotic; cytostatic; vulnery; hyaluronic acid antagonist;    |
| KW  | HA binding inhibitor; multiple sclerosis; restenosis; diabetes mellitus;  |
| KW  | neurological disease; Alzheimer's disease; Parkinson's disease; cancer;   |
| KW  | arthritis; inflammatory joint disease; osteoporosis; atherosclerosis;     |
| KW  | inflammatory dermatosis; inflammatory bowel disease; metastasis;          |
| KW  | inflammatory disease; respiratory distress syndrome; wound healing;       |
| KW  | RHAMM binding protein; RABP.  |
| XX  |   |
| XX  | Mus sp.   |
| XX  |   |
| XX  | WO2003033535-A2.  |
| PN  |   |
| XX  | 24-APR-2003.  |
| PD  |   |
| XX  | 15-OCT-2002; 2002WO-CA001563.   |
| XX  |   |
| PP  | 15-OCT-2001; 2001US-00978309.   |
| PR  |   |
| XX  | (TRAN-) TRANSITION THERAPEUTICS INC.                                      |
| PA  |   |
| XX  |   |
| PI  | Cruz T, Pastrak A, Turley EA;   |
| FI  |   |
| XX  | WPI; 2003-372323/35.  |
| DR  |   |
| XX  | New P16, P-16d, human and murine P32, S3, S7, and V2 polypeptides, useful |
| PT  | for treating e.g. multiple sclerosis, restenosis, diabetes mellitus,      |
| PT  | Alzheimer's disease, Parkinson's disease, arthritis, cancer and           |
| PT  | metastasis.   |
| XX  |   |
| PS  | Claim 1; Page 27; 215pp; English.   |
| XX  |   |
| CC  | The present invention describes a polypeptide (I) which binds to          |
| CC  | hyaluronic acid (HA) and thereby inhibit the binding of HA to RHAMM       |
| CC  | (receptor hyaluronic acid mediated motility), comprising an amino acid    |
| CC  | sequence of P16, P-16d, human P32, murine P32, murine S3, human S3,       |
| CC  | murine S7, human S7, murine V2 and human V2. Also described: (1) a        |
| CC  | pharmaceutical composition comprising (1); (2) an antibody which binds to |
| CC  | (1) or to a polypeptide comprising murine V3 or human V3 amino acid       |
| CC  | sequence; (3) a vaccine composition for the treatment of multiple         |
| CC  | sclerosis, restenosis or diabetes mellitus, comprising an antigen for the |
| CC  | antibody of (2); and (4) a method for treating multiple sclerosis, or     |
| CC  | restenosis or diabetes mellitus, by administering (1), human or murine    |
| CC  | V3, or an antibody to these polypeptides. (1) has neuroprotective,        |
| CC  | vasotropic, antidiabetic, nootropic, antiparkinsonian, antiarthritic,     |
| CC  | antiinflammatory, osteopathic, dermatological, antiarteriosclerotic,      |
| CC  | cytostatic and vulnery activities, and can be used in vaccines and as     |
| CC  | an HA antagonist. The polypeptides and the antibodies to these            |
| CC  | polypeptides are useful for treating multiple sclerosis, restenosis or    |
| CC  | diabetes mellitus, as well as neurological diseases such as Alzheimer's   |
| CC  | disease, Parkinson's disease, arthritis and other inflammatory joint      |
| CC  | diseases, osteoporosis, inflammatory dermatosis, inflammatory bowel       |
| CC  | disease and other inflammatory diseases, atherosclerosis and related      |
| CC  | diseases, cancer and metastasis, and chronic and acute respiratory        |
| CC  | distress syndrome. The polypeptides may also be used to promote wound     |
| CC  | healing. ABR56968 to ABR57022 and ACC79533 to ACC79537 represent          |
| CC  | sequences used in the exemplification of the present invention            |
| XX  |   |
| XX  | Sequence 477 AA;  |
| SQ  |   |
| Query Match 7.4%; Score 111; DB 6; Length 477;                  |   |
| Best Local Similarity 20.5%; Pred. No. 1.2;                     |   |
| Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15 |   |
| QY  | 1 MTEIVADKTVVEVKNAIETADGALDLYNKYLDVIVPQTFDETKEISRFQBSYQAASV 60            |
| DB  | 99 LNLNLEKEVELEKHIAHAQAAILTAQEKYNDTA---QSLRDVVTQALESVQEKYNDTAQS 155       |
| QY  | 61 LVGDIKTLMLDSQDKYFEATQTVVEWCGVATQALLAAVILLFDENYKKAQKDILIKVL 120         |
| DB  | 156 L-RDVTVAQLESQEKYNDTAQSLRD---VTQALESQ-----EKYND-TAQLRDVTAQ-L 205       |
| QY  | 121 DGIITKLNBAQKSL-L-VSSQSFNNAAGKLLAL-DSQLTN-DFSFKSYFQSDVKIRKEA 177       |

|  |   |  |   |
|--|---|--|---|
| human S7, murine V2 and human V2, all derived from the protein RHAMM (Receptor hyaluronate mediated motility) a cell surface target protein that is required for activation of the AP-1 (activating protein-1) pathway on response to injury. Also included are an antibody that binds to the polypeptide cited above, a pharmaceutical or vaccine compositions for treating an inflammatory neurological disorder or diabetes mellitus (comprising the above amino acid sequence or an antigen for the antibody cited above) and treating an inflammatory bowel disease, disorder, arthritis, inflammatory dermatosis, inflammatory lung disease, cancer, kidney fibrosis, inflammatory lung disease, obesity, lupus, cardiovascular disease, diabetes mellitus or wounds (comprising administering to a patient the above polypeptide or antibody. In the above method, the inflammatory neurological disorder is multiple sclerosis, Parkinson's disease or Alzheimer's disease. The polypeptide or antibody may also be used for diagnosing or preventing the above diseases. The present sequence is a mouse RHAMM active protein fragment. |   | human S7, murine V2 and human V2, all derived from the protein RHAMM (Receptor hyaluronate mediated motility) a cell surface target protein that is required for activation of the AP-1 (activating protein-1) pathway on response to injury. Also included are an antibody that binds to the polypeptide cited above, a pharmaceutical or vaccine compositions for treating an inflammatory neurological disorder or diabetes mellitus (comprising the above amino acid sequence or an antigen for the antibody cited above) and treating an inflammatory bowel disease, disorder, arthritis, inflammatory dermatosis, inflammatory lung disease, cancer, kidney fibrosis, inflammatory lung disease, obesity, lupus, cardiovascular disease, diabetes mellitus or wounds (comprising administering to a patient the above polypeptide or antibody. In the above method, the inflammatory neurological disorder is multiple sclerosis, Parkinson's disease or Alzheimer's disease. The polypeptide or antibody may also be used for diagnosing or preventing the above diseases. The present sequence is a mouse RHAMM active protein fragment. |   |
| CC   | Query Match   | 7.4%;  | Score 111; DB 7; Length 477;            |
| CC   | Best Local Similarity   | 20.5%;   | Pred. No. 1.2; Indels 62; Gaps 15;      |
| CC   | Matches   | 60; Conservative   | 63; Mismatches 108; Indels 62; Gaps 15; |
| CC   | 1 MTEIVADKTVVVKNAIETADGALDLYNKYLQVVPWQTFDETIKELSRFKQYSQAASV 60  |  |   |
| CC   | 99 LDNLLREKEVELEKHAHAQAIIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155   |  |   |
| CC   | 61 LVGDIKTLMDSDQKYFEATQVYEWGCVATQLLAAYILLDFYNEKKAQAOKDILIKVL 120  |  |   |
| CC   | 156 L-RDVTAQLESVQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQLRDVTAQ-L 205  |  |   |
| CC   | 121 DDGITKLENAQKSL-L-VSSQSFNNAGKLLAL-DSQLTN-DFSEKSSYFQSQVDKIRKEA 177  |  |   |
| CC   | 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLENLTQEKVAMAEKSVEDVQQOI 265  |  |   |
| CC   | 178 YAGAAAGVAVGPFGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNF-----TTLS 226   |  |   |
| CC   | 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEEIKETTSFLEKITDLK 308   |  |   |
| CC   | 227 NTVKQANKDI-----DAAKLKTETIAAIGEIKTETETTRFYVDYDDL 269   |  |   |
| CC   | 309 NQLRQODEDFRKLQEKKGKRTAEKENVMTELT-----MEINKWRLLYEEL 353  |  |   |
| CC   | RESULT 37   |  |   |
| CC   | AD020468  |  |   |
| CC   | ID ADC02468 standard; protein; 477 AA.  |  |   |
| CC   | AC ADC02468;  |  |   |
| CC   | 18-DEC-2003 (first entry)   |  |   |
| CC   | Mouse RHAMM V2 protein #1.  |  |   |
| CC   | RHAMM; Receptor hyaluronate mediated motility;  |  |   |
| CC   | cell surface target protein; AP-1; activating protein-1;  |  |   |
| CC   | response-to-injury; vaccine; inflammatory neurological disorder;  |  |   |
| CC   | diabetes mellitus; arthritis; inflammatory dermatosis;  |  |   |
| CC   | inflammatory bowel disease; cancer; kidney fibrosis;  |  |   |
| CC   | inflammatory lung disease; obesity; lupus; cardiovascular disease; wound;   |  |   |
| CC   | multiple sclerosis; Parkinson's disease; Alzheimer's disease;   |  |   |
| CC   | alpha helical peptide; mouse.   |  |   |
| CC   | Mus sp.   |  |   |
| CC   | US2003100490-A1.  |  |   |
| CC   | 29-MAY-2003.  |  |   |
| CC   | 15-OCT-2001; 2001US-00978309.   |  |   |
| CC   | 01-APR-1999; 99US-0127457P.   |  |   |
| CC   | 03-APR-2000; 2000US-00541522.   |  |   |
| CC   | 05-OCT-2000; 2000US-00685010.   |  |   |
| CC   | (CRUZ/) CRUZ T.   |  |   |
| CC   | (PAST/) PASTRAK A.  |  |   |
| CC   | (TURL/) TURLEY E A.   |  |   |
| CC   | Cruz T, Pastrak A, Turley EA;   |  |   |
| CC   | WPI; 2003-755209/71.  |  |   |
| CC   | New polypeptide for diagnosing, preventing or treating inflammatory neurological disorders, cancer, obesity, lupus or diabetes mellitus, comprises an amino acid sequence of P16, P32, or murine or human S3, S7 or V2. |  |   |
| CC   | Claim 1; SEQ ID NO 77; 110pp; English.  |  |   |
| CC   | The invention relates to a polypeptide comprising an amino acid sequence selected from the sequences of P16, P32, murine S3, human S3, murine S7,   |  |   |



SQ Sequence 794 AA;  
Query Match 7.4%; Score 111; DB 5; Length 794;  
Best Local Similarity 20.5%; Pred. No. 2.4;  
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;  
QY 1 MTEIVADKTEVEVKNIAETADGALDLYNKYLDQVIPHQTDETIKELSRKQESQAASV 60  
DB 416 LDNLREKEVELEKHIAHAQAIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 472  
QY 61 LVGDIKTLMDSDQKYEATQTVVWCGVATQLLAAVILLFDEYNEKKASQAQDILIKVL 120  
DB 473 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQLSRDVTQA-L 522  
QY 121 DDGITKLINEAQKSL-L-VSSQSFNNASGKLAL-DSQLTN-DFSEKSSYFOSQVDKIRKEA 177  
DB 523 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLNTLQEKVAMAEKSVEDVQQOI 582  
QY 178 YAGAAAGVAVGFFGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQFF---TTLS 226  
DB 583 LTAESTNQ-----EYA-----RMVQDQNRSTLKEEIKETISFLEKITDLK 625  
QY 227 NTVKQANKDI-----DAAKLIKLTETIAALGEIKETETTRFYVDYDDL 269  
DB 626 NQLRQODEDFRKLQLEKGRPTAEKENVMTLT-----MEINKRLLLYEEL 670  
RESULT 40  
ID ABU43682  
XX ABU43682 standard; protein; 938 AA.  
AC ABU43682;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #29209.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Staphylococcus haemolyticus.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA47552.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 71606; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway; (8)  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 938 AA;  
Query Match 7.4%; Score 111; DB 6; Length 938;  
Best Local Similarity 18.6%; Pred. No. 3;  
Matches 60; Conservative 64; Mismatches 110; Indels 88; Gaps 11;  
QY 9 TVEVKNIAETADGALDLYNKYLDQVIPHQTDETIKELSRKQESQAASVVGDIKTL 68  
DB 189 TINKIKNAVVAHDSLPQINKIADRI---EYLNDHODDLDKYANQFR-----ALGNKGD 240  
QY 69 LMDSDQKYEATQTVVWCGVATQLLAAVILLFDEY-----NEKKASQAQ--- 113  
DB 241 ILDRQOKLNDVNAI-----PSLNEKAKLILALNEYMPNIEKLDDVASNDIPAQFPKINR 295  
QY 114 --DILIKVLDDGTTKLINEAQKSLLSQ---SPNNASGKLALDLSQTLNDFSEKSS--YF 166  
DB 296 GVDIASEGFDLANTRLNDACQYLTSQAQORVGDYQEAAGRAQEVNNOANSALRQOQSGLP 355  
QY 167 QSQVDKIRKEAYAGAAAGVAVGFFGLIISYSIAAGVVEGKLIPELKNKLSKVQNFPTLS 226  
DB 356 QYQIQLSTD-----NSQDTVNDNQIVSN 379  
QY 227 NTVKQANKDIDAAKLIKLT-----TEIAAIGEIKETETTRFYVDYDDLMLSL-- 273  
DB 380 NDVKSMNSALAEALLTSSNSDNOAKATQSDIKALKADISYGVIGSNRPTEFNDMLRNKLT 439  
QY 274 -LKEAAK---KMINTCNEYOKR 291  
DB 440 RLENSKSNQQLIDVLKELEKR 461  
RESULT 41  
ID ABB60789  
XX ABB60789 standard; protein; 1013 AA.  
XX  
AC ABB60789;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 9159.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
FN



```

QY 266 -----YDMLSLKKEAAKWMINTCNEYOKR 291
Db 460 KTKQLEIDSKLGFPTKIHQLRDESELSIQKNFNQ 497

RESULT 43
ID ABP26411 standard; protein; 304 AA.
XX AC ABP26411;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 1998.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus agalactiae.
XX FN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN67042.
XX PT New Streptococcus protein for the treatment or prevention of infection or
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX PT detecting a compound that binds to the protein.
XX PS Claim 1; Page 3350; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins
XX SQ Sequence 304 AA;
XX Query Match 7.3%; Score 110.5; DB 5; Length 304;
XX Best Local Similarity 21.6%; Pred. No. 0.71;
XX Matches 69; Conservative 57; Mismatches 126; Indels 67; Gaps 16;
QY 10 VEVKNAIETADGALDLYNKLQDVPWQTFETIKLSRFKQ-----EYSQASVLIV 62
Db 4 VSVSSVGTQASTVATSMFSRV-----SALNDATKLSFAEATLQGTATYNAKSYAT 56

```

63 GDIKTLMDSDQKYFEATQTVYWCQVATQLLAAAYILLFDEYN-----EKASAKODIL 116  
 57 GTLTPMLQG---MILFSETLSEKC---TELQTLVYVIGCDEDLSDVWLESKLASDRSL 109  
 117 IKVLDDGITYKLN---BAQSLIVSSQS-FNNASGKLLALDSQI--TNDFSEKSSYFQSQV 170  
 110 -KIAEALLEHLNDPPEPSKSAISSTKSNITKUKRIKSNQKLLDNJNEFNAHSATVFADI 168  
 171 DKIR---KEAYAGAAAGV-----VAGPFGILLISYSIAAGVV-----EGKLIPELKNK 214  
 169 SNAQSTVNOALAAVSTGFGSYNSKTCGAFCKPTSGQMEWTKVKKNWKEREDAKABELKSK 228  
 215 LKSVQNFPTTLGNTVQKANDIDAKL-KLTTEIAAIGIKETETETTRFYVDYDDL--- 270  
 229 KAEESKASKIENTTKKSNVSVDBKKLKAANEAYKLGEEKDT-----YESIISGL 280  
 271 ----LSLKEAAK-KMINT 284  
 281 SNASALLKEVAKSKLTD 299

RESULT 44  
 ADK99510  
 ID ADK99510 standard; protein; 304 AA.  
 XX AC ADK99510;  
 XX DT 20-MAY-2004 (first entry)  
 XX DE Streptococcus agalactiae ORF SAG1030-related protein 1.  
 XX KW immunogenic composition; group B Streptococcus; GBS; antibacterial;  
 XX KW streptococcal infection; vaccine; SAG.  
 XX OS Streptococcus agalactiae 2603V/R.  
 XX FN WO2004018646-A2.  
 XX PD 04-MAR-2004.  
 XX PF 26-AUG-2003; 2003WO-US026827.  
 XX PR 26-AUG-2002; 2002US-0406237P.  
 XX PR 27-AUG-2002; 2002US-0406676P.  
 XX PR 28-AUG-2002; 2002US-0406757P.  
 XX PA (CHIR ) CHIRON CORP.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Tettelin H, Masignani V;  
 XX DR WPI; 2004-248071/23.  
 XX PT Immunogenic composition useful as a vaccine for treating or preventing  
 XX PT streptococcal infections, comprises group B Streptococcus polypeptides.  
 XX PS Claim 10; SEQ ID NO 6104; 1194pp; English.  
 XX CC The invention relates to a novel immunogenic composition comprising a  
 XX CC combination of 2-5 group B Streptococcus (GBS) polypeptides. Each  
 XX CC polypeptide is encoded by a GBS polynucleotide sequence which is  
 XX CC homologous to a polynucleotide sequence of group A Streptococcus (GAS),  
 XX CC Streptococcus pneumoniae and/or least one other GBS serotype. The  
 XX CC composition of the invention demonstrates antibacterial activity whilst  
 XX CC the polypeptides and polynucleotides may be useful in assays to diagnose  
 XX CC and identify streptococcal infections or for identifying, screening and  
 XX CC developing vaccines and other treatments for streptococcal infections.  
 XX CC The current sequence is that of a Streptococcus agalactiae ORF SAG  
 XX CC protein of the invention.  
 XX SQ Sequence 304 AA;





CC Streptococcus pneumoniae and/or least one other GBS serotype. The  
CC composition of the invention demonstrates antibacterial activity whilst  
CC the polypeptides and polynucleotides may be useful in assays to diagnose  
CC and identify streptococcal infections or for identifying, screening and  
CC developing vaccines and other treatments for streptococcal infections.  
CC The current sequence is that of a Streptococcus agalactiae ORF SAG  
CC protein of the invention.  
XX  
SQ Sequence 278 AA;  
  
Query Match 7.3%; Score 110; DB 8; Length 278;  
Best Local Similarity 22.9%; Pred. No. 0.69;  
Matches 66; Conservative 50; Mismatches 112; Indels 60; Gaps 15;  
  
QY 41 DETIKELSRFKQ-----EYSQAASVLVVDIKTLMDSDQKYPEATQTVYVCGVATQL 93  
DB 2 NDAITKLSSFAEAAATLQGTAYNSAKSVATGTLTPMLOG-----MLFSETISEKC---TEL 54  
  
QY 94 LAAYILLFDEYN-----EKKASAKDILIKVLDDGITKLN---EAKSLLVSSQS-FNN 143  
DB 55 QTLVYVICGDELDLSVLESKLADRSAL-KIAALLHLEHNDPEPSKSAISSTKSNIKK 113  
  
QY 144 ASGKLLALDSQL--TNDPSEKSSYFQSQVDKIR---KEAVAGAAAGV-----VAGPGLI 193  
DB 114 LKKRIKSNQKLDNLNEFNHSAATVPADISNAQSTVWQALAAVSTGFGSYNSKTFGAFGKP 173  
  
QY 194 ISYSTAAGV-----EGKLIPELKNKLSVQNFVFTLSNTVKQANKDIDAACL-KLT 244  
DB 174 TSGQWMTKTVKKNWKEREDAKAEELSKAEBSKKASKIENTTKNSVSDVKKLLKAA 233  
  
QY 245 TEIAAIGELKTETETRFVVDYDILM-----LSLLKEAAK-KMINT 284  
DB 234 NEAYKLGEIKKDT-----YESIISGLSNASAAALLKEVAKSKLTD 273

RESULT 47  
ABU23940  
ID ABU23940 standard; protein; 1163 AA.  
XX  
AC ABU23940;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #9467.  
XX  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
XX Clostridium acetobutylicum.  
XX  
XX WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX  
PR 06-SEP-2001; 2001US-00948993.  
XX  
PR 25-OCT-2001; 2001US-0342923P.  
XX  
PR 08-FEB-2002; 2002US-00072851.  
XX  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
XX  
XX N-PSDB; ACA27810.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX

PS  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1163 AA;

Query Match 7.3%; Score 110; DB 6; Length 1163;  
Best Local Similarity 17.9%; Pred. No. 4.9;  
Matches 73; Conservative 69; Mismatches 114; Indels 152; Gaps 14;  
  
QY 1 MTEIVADKTVVYVKNAIETADG-----ALDLYNKYLDQVTPWTFDETIK----- 45  
DB 109 IVDITGDE-VEVLEEGAKSVNEXCOBIIIGLSLDDFTRTV--VLPOGKFSEFLKLEGKERR 165  
  
QY 46 -----ELSRFQSVSOAASVLVGDIK-----TLLMDSQ 73  
DB 166 NMLERLNLQYGDLSFLKARKIRKEREKENVLVGELKGYENINEDVLKERRELLKENN 225  
  
QY 74 DKYFEATQTVYVWCGVATQLAAAYILLFDSYN-----EKKASAKDILIKV-- 119  
DB 226 DFFNEASK-----EYLKAEERYNEGKEVWGLQIBIEEKRVKDKLMKKODE 271  
  
QY 120 --LDDGITKLENAQKSLVSSQSPFNASGKLLALDSQLTN-----DFSEKSS 164  
DB 272 IDLKEKEARLGESSKVKPYDINVENTLKQIDILKEQILSRENTMKAIISLEKEDMEKLS 331  
  
QY 165 YFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYSIAGVVEGKLIPELKNLKSQVNFPTT 224  
DB 332 IAKONKEK-----ALPKFMIKHIIILDAIKEKDLNKLKLEKRLQGIEX 377  
  
QY 225 LS-----NTVKQANKDIDAOKLJTEAATGEIKTETE----- 258  
DB 378 LSLEASNKEELIKQNTKIDISLTLKIQNLESKIDNLKVPPEYKKNKINEGIFLLRNYDEKL 437  
  
QY 259 -----TTRFVVDYD-----DLMSLLKEAAKKMINTCFYQKR 291  
DB 438 KHKNLGLDCKDFQVDFEKAASKEMLFNKLSEERSKL----DTYTKK 481

RESULT 48  
ADE56031  
ID ADE56031 standard; protein; 3187 AA.  
XX



AC ADE56031;  
 XX 29-JAN-2004 (first entry)  
 DT Rat Protein BAA05026, SEQ ID NO 1870.  
 XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
 XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
 XX Rattus norvegicus.  
 XX WO2003016475-A2.  
 XX 27-FEB-2003.  
 XX 14-AUG-2002; 2002WO-US025765.  
 XX 14-AUG-2001; 2001US-0312147P.  
 XX 01-NOV-2001; 2001US-0346382P.  
 XX 26-NOV-2001; 2001US-0333347P.  
 XX (GEHO ) GEN HOSPITAL CORP.  
 XX (FARB ) BAYER AG.  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 XX WPI; 2003-268312/26.  
 XX GENBANK; BAA05026.  
 XX New composition comprising two or more isolated polypeptides, useful for  
 XX preparing a medicament for treating pain in an animal.  
 XX Claim 1; Page; 1017pp; English.  
 XX The invention discloses a composition comprising two or more isolated rat  
 XX or human polynucleotides or a polynucleotide which represents a fragment,  
 XX derivative or allelic variation of the nucleic acid sequence. Also  
 XX claimed are a vector comprising the novel polynucleotide, a host cell  
 XX comprising the vector, a method for identifying a nucleotide sequence  
 XX which is differentially regulated in an animal subjected to pain and a  
 XX kit to perform the method, an array, a method for identifying an agent  
 XX that increases or decreases the expression of the polynucleotide sequence  
 XX that is differentially expressed in neuronal tissue of a first animal  
 XX subjected to pain, a method for identifying a compound which regulates  
 XX the expression of a polynucleotide sequence which is differentially  
 XX expressed in an animal subjected to pain, a method for identifying a  
 XX compound that regulates the activity of one or more of the  
 XX polynucleotides, a method for producing a pharmaceutical composition, a  
 XX method for identifying a compound or small molecule that regulates the  
 XX activity in an animal of one or more of the polypeptides given in the  
 XX specification, a method for identifying a compound useful in treating  
 XX pain and a pharmaceutical composition comprising the one or more  
 XX polypeptides or their antibodies. The polynucleotide or the compound that  
 XX modulates its activity is useful for preparing a medicament for treating  
 XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 XX therapy). The sequence presented is a rat protein (shown in Table 2 of  
 XX the specification) which is differentially expressed during pain. Note:  
 XX The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic form directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 3187 AA;  
 Query Match 7.3%; Score 110; DB 7; Length 3187;  
 Best Local Similarity 20.8%; Pred. No. 19;  
 Matches 80; Conservativity 65; Mismatches 114; Indels 126; Gaps 18;  
 QY 1 MTEIVADKTVB---VVRNA-IETADGALDLYNKYLDQVPMQTFDEIKELSRFKYQSQ 56  
 Db 1725 VTEAVGKSGQSDLSSENAKLEDAEATL-LANSAPGVGS--ETP-SSHDDINNYLQLDQ 1780  
 QY 57 AASVLVGDITKLMDSDQKYFEATQTVEWCGVATQLLAAILLFDEYNEKKA-----SAQ 112

Db 1781 ----LKGRIAELEMEKO-KDRELSQTL-----ENEKNALLTQISA 1815  
 QY 113 KDILIKVLDGDIITKLN----EAQKSLIVSSQSFNNASGKLALDSQLTNDSEKS----S 164  
 DE 1816 KDSSELKLEBEVAKINMLNQIQEELSRTVKLTABEEXKDDLEERLWNLQLAELNGSIGN 1875  
 QY 165 YFQSQVD-----KIRKEAYAGAAAGVV 186  
 Db 1876 YYQDVTDQAQIKNEQSEMQNLKRCVSELEEEKQOLVKEKTKVSEIRKE-YMEKIQGAQ 1934  
 QY 187 AGPFGLIISYSIAAGVVEGKLIPELKNKLSVQ-----NFFTTLNNTVK-----QA 232  
 Dp 1935 KPGSKIHAKELQ-----ELLKEQOEVLQKDCIRYLGRISALEKTVKALEFVHTES 1988  
 QY 233 NKDIDAAK-----LKLTTETIAAIGBIKTETETTRFYVDYDDLML----- 271  
 Db 1989 QKOLDATKGNLAQAVEHHKKAQAELSFKILLDDTQSEAAARVLADNLKELQSNKESI 2048  
 QY 272 -SLLKEAAKQWINTCNEYQKRGKK 295  
 Db 2049 KSQIKQKQEDLLRLLEQAEEKHKE 2073  
 RESULT 49  
 ADE56035  
 ID ADE56035 standard; protein; 3187 AA.  
 XX AC ADE56035;  
 XX DT 29-JAN-2004 (first entry)  
 XX DE Rat Protein BAA05026, SEQ ID NO 1874.  
 XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
 XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
 XX Rattus norvegicus.  
 XX WO2003016475-A2.  
 XX 27-FEB-2003.  
 XX 14-AUG-2002; 2002WO-US025765.  
 XX 14-AUG-2001; 2001US-0312147P.  
 XX 01-NOV-2001; 2001US-0346382P.  
 XX 26-NOV-2001; 2001US-0333347P.  
 XX (GEHO ) GEN HOSPITAL CORP.  
 XX (FARB ) BAYER AG.  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 XX WPI; 2003-268312/26.  
 XX GENBANK; BAA05026.  
 XX New composition comprising two or more isolated polypeptides, useful for  
 XX preparing a medicament for treating pain in an animal.  
 XX Claim 1; Page; 1017pp; English.  
 XX The invention discloses a composition comprising two or more isolated rat  
 XX or human polynucleotides or a polynucleotide which represents a fragment,  
 XX derivative or allelic variation of the nucleic acid sequence. Also  
 XX claimed are a vector comprising the novel polynucleotide, a host cell  
 XX comprising the vector, a method for identifying a nucleotide sequence  
 XX which is differentially regulated in an animal subjected to pain and a  
 XX kit to perform the method, an array, a method for identifying an agent  
 XX that increases or decreases the expression of the polynucleotide sequence  
 XX that is differentially expressed in neuronal tissue of a first animal  
 XX subjected to pain, a method for identifying a compound which regulates  
 XX the expression of a polynucleotide sequence which is differentially  
 XX expressed in an animal subjected to pain, a method for identifying a  
 XX compound that regulates the activity of one or more of the  
 XX polynucleotides, a method for producing a pharmaceutical composition, a  
 XX method for identifying a compound or small molecule that regulates the  
 XX activity in an animal of one or more of the polypeptides given in the  
 XX specification, a method for identifying a compound useful in treating  
 XX pain and a pharmaceutical composition comprising the one or more  
 XX polypeptides or their antibodies. The polynucleotide or the compound that  
 XX modulates its activity is useful for preparing a medicament for treating  
 XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 XX therapy). The sequence presented is a rat protein (shown in Table 2 of  
 XX the specification) which is differentially expressed during pain. Note:  
 XX The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic form directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences.

expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a rat protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 3187 AA;

Query Match 7.3%; Score 110; DB 7; Length 3187;

Best Local Similarity 20.8%; Pred. No. 19; Mismatches 80; Conservative 65; Indels 126; Gaps 18;

QY 1 MTEIVADKTV---VVKNA-IETADGALDLYNKYLDQVIMQTFDETIKLSRFKQVYSQ 56  
Db 1725 VTEAVVGKSEQODLSSENAKLEDAEATL-LANSAPKGV---ETP-SSHDDINNYLQQLDQ 1780  
QY 57 AASVLVGDIKTLMSQKPYFATQVYVWCVATQALLAAILFDEYNEKKA-----SAQ 112  
Db 1781 -----LKGRFAELEMEKQ-KDRELSQTL-----ENEKNALLTQISA 1815  
QY 113 KDILIKVLDDGTTKLN-----EAQKSLVSSQSFNNASGKLALDSQLTNDFSEKS-----S 164  
Db 1816 KQSEKLLLEEYAKINMLNQIQEELSRVTKLTAEEBKDDLERLMNLQAEELNGSIGN 1875  
QY 165 YFQSQVD-----KIRKEAYAGAAAGV 186  
Db 1876 YQDVTDAQIKNEQLESEMNLRKCVSELEEKQOLVKEKTVSEIRKE-YMEKIQAQ 1934  
QY 187 AGPFGLIISYSAAGVVEGKLIPELKNKLSVQ-----NPFITLSNTVK-----QA 232  
Db 1935 KGGFSKIHAKELQ-----ELKEKQEVKQKQKDCIRYLGRIKSALEKTVKALEFVHTES 1988  
QY 233 NKDIDAAK-----LKLTEIAAIGEIKETETTRFYVDYDML----- 271  
Db 1989 QKDLATKGNLAQAVEHHKKAQAEELSSFKILLDDTQSEAAARVLADNKLKKELOSNKESI 2048  
QY 272 -SLLKEAAKMTNCNEYOKRHKK 295  
Db 2049 KSQIKQKDEDLRLRLEQAEEKIRKE 2073

RESULT 50

AAG23551  
ID AAG23551 standard; protein; 366 AA.

XX AC AAG23551;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 26901.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX EP1033405-A2.

XX FN 06-SEP-2000.

XX PD 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 22-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139452P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140981P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.

PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 17-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147200P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.

PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 04-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 7.3%; Score 109.5; DB 3; Length 366;

Best Local Similarity 22.2%; Pred. No. 1.1; Mismatches 41; Indels 95; Gaps 15;  
Matches 76; Conservative

Qy 7 DKTVEVVKNAIETADGALDLYNKYLDQVDPWT-----FDETIKELSRFKQBYSQ 56  
Db 52 DAHMELSKHLVEVTQGVADFIETEDDV--WDNQAKYVLVAYFENTKKTLEIFKTIENC 109  
Qy 57 AASVLVGDIKTLMLDSODKYFEATQTVYVWCGVATQLLAAYILLDFBYNEKKAQAQDI- 115  
Db 110 VENAEMGQL-----LITREALBF-----EKESAEDVG 137  
Qy 116 -----LIKVLDDGITKLNKAEQKSLVSSQSFNNASGKLAL-----DSQLTNDPSEK 162  
Db 138 GKXKKYEKTLIED-----LKSPKMGDPDP---GKVLTTQFELIKKQESLLEEVSET 186  
Qy 163 SSVFQSOVDKIRKEAYA-----GAAAGVAGPFGLIISYIAAGVVEGKLIPELKNKLKS 217  
Db 187 KKKIQDEITNLEKTKTLITNVVFGAFAVAVASIALIATGVGAAGFALAAPL--LAAG 244  
Qy 218 VQNFPTLSNTVQANKIDAAKUKLTETIAAIGEIKTE---TETTRFYD-YDDLMLSL 273  
Db 245 WAGVYTTLDKKDALKQLEGLKKVEIEESVEKGIKTNEEATETVSLVDGLEDRINKM 304  
Qy 274 LK-----EAAKKMINTCNEYOKRHGK--KTLFEVPE 302  
Db 305 LKLVDNAIDHEDNEAATRIVLT--QISKVKEKLTUKKITEVGE 344

Search completed: January 5, 2005, 10:56:33  
Job time : 68.2911 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:43:53 ; Search time 55.3174 Seconds  
(without alignments)  
3151.602 Million cell updates/sec

Title: US-09-993-292B-28  
Perfect score: 1508  
Sequence: 1 MTEIVADKTVVKNNAIETA.....TCNEYQKRHKKTLFEVPEV 303

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database : UniProt 02.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID        | Description          |
|------------|-------|-------------|--------|--------------|----------------------|
| 1          | 1503  | 99.7        | 302    | 1 HLYE_ECOLI | P77335 escherichia   |
| 2          | 1488  | 98.7        | 302    | 1 HLYE_ECO57 | Q9reb3 escherichia   |
| 3          | 1395  | 92.5        | 302    | 1 HLYE_SALTI | Q8z727 salmonella    |
| 4          | 1376  | 91.2        | 302    | 1 HLYE_SALPA | Q93rr6 salmonella    |
| 5          | 1133  | 75.1        | 300    | 2 Q9X288     | Q9x288 escherichia   |
| 6          | 563   | 37.3        | 113    | 1 HLEL_SHIFL | Q9rcr3 ehigella fl   |
| 7          | 437   | 29.0        | 93     | 1 HLEL_ECOL6 | Q8fi27 escherichia   |
| 8          | 133.5 | 8.9         | 584    | 1 SYR_SYNY3  | Q55486 synecocyst    |
| 9          | 129.5 | 8.6         | 1906   | 2 Q6C359     | Q6c359 yarrowia li   |
| 10         | 128   | 8.5         | 495    | 2 P71497     | P71497 mycoplasma    |
| 11         | 126   | 8.4         | 652    | 2 Q7CMF0     | Q7cmf0 bacillus an   |
| 12         | 126   | 8.4         | 652    | 2 Q9X360     | Q9x360 bacillus an   |
| 13         | 126   | 8.4         | 652    | 2 AAT28865   | Aat28865 bacillus    |
| 14         | 126   | 8.4         | 979    | 2 Q9EP71     | Q9ep71 mus musculu   |
| 15         | 126   | 8.4         | 992    | 2 Q6ZPT6     | Q6zpt6 mus musculu   |
| 16         | 126   | 8.4         | 992    | 2 BAC98143   | Bac98143 mus muscu   |
| 17         | 125   | 8.3         | 504    | 2 Q8I752     | Q8i752 plasmodium    |
| 18         | 125   | 8.3         | 2749   | 2 Q7REY3     | Q7rey3 plasmodium    |
| 19         | 124.5 | 8.3         | 1081   | 2 Q73KC5     | Q73kc5 treponema d   |
| 20         | 124.5 | 8.3         | 1081   | 2 AAS12812   | Aas12812 treponema   |
| 21         | 123.5 | 8.2         | 465    | 2 Q73AJ3     | Q73aj3 bacillus ce   |
| 22         | 123.5 | 8.2         | 465    | 2 AAS40708   | Aas40708 bacillus    |
| 23         | 123.5 | 8.2         | 713    | 2 Q6MT03     | Q6mt03 mycoplasma    |
| 24         | 123.5 | 8.2         | 713    | 2 CAE77235   | Caet77235 mycoplasma |
| 25         | 123   | 8.2         | 541    | 2 Q74DE2     | Q74de2 geobacter s   |
| 26         | 123   | 8.2         | 541    | 2 AAR34750   | Aar34750 geobacter   |
| 27         | 123   | 8.2         | 649    | 2 Q9LVQ4     | Q9lvq4 arabidopsis   |
| 28         | 123   | 8.2         | 1189   | 2 Q8CPT0     | Q8cpt0 staphylococ   |
| 29         | 123   | 8.2         | 2752   | 2 Q9BJY0     | Q9bjy0 plasmodium    |
| 30         | 122.5 | 8.1         | 927    | 2 Q9LX69     | Q9lx69 listeria mo   |
| 31         | 122.5 | 8.1         | 927    | 2 AAT05096   | Aat05096 listeria    |

## ALIGNMENTS

## RESULT 1

HLYE\_ECOLI  
ID HLYE\_ECOLI STANDARD; PRT; 302 AA.  
AC P77335; Q47276; Q8VU70; Q9R3G4;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 29-MAR-2004 (Rel. 43, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Hemolysin E, chromosomal (Hemolysin-inducing protein) (Silent  
DE Hemolysin sheA) (Cytotoxin clyA) (Latent pore-forming 34 kDa  
DE haemolysin).  
GN Names: hlyE; Synonyms: clyA, sheA, hpx; OrderedLocusNames=b1182;  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / XL1-BLUE;  
RA McNamara P.J., Iandolo J.J., Uhlich G.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA del Castillo F.J., Leal S.C., Moreno F., del Castillo I.;  
RL "The Escherichia coli K-12 sheA gene encodes a 34-kDa secreted  
RT haemolysin".  
RL Mol. Microbiol. 25:107-115(1997).  
RN [3]  
RP SEQUENCE FROM N.A., SEQUENCE OF 1-12, SUBCELLULAR LOCATION, AND  
RP INDUCTION.  
RC STRAIN=K12;  
RX MEDLINE=99157562; PubMed=10027972;  
RA Ludwig A., Bauer S., Benz R., Bergmann B., Goebel W.;  
RT "Analysis of the SlyA-controlled expression, subcellular localization  
RT and pore-forming activity of a 34 kDa haemolysin (ClyA) from  
RT Escherichia coli K-12".  
RL Mol. Microbiol. 31:557-567(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3030-2;  
RA Xing J., Fernandez S.V., Kapur V., Barletta R.G., Moxley R.A.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

32 122.5 8.1 4007 2 Q7SHZ4  
33 122 8.1 852 2 Q8I754  
34 122 8.1 1496 2 Q9SZK7  
35 121.5 8.1 1081 2 Q7RFQ7  
36 121 8.0 1116 2 Q7RAW8  
37 120.5 8.0 961 2 Q8N592  
38 120.5 8.0 962 1 VDP\_HUMAN  
39 120.5 8.0 1611 2 Q7RD43  
40 120 8.0 1012 2 Q6GNE7  
41 120 8.0 1102 1 MYSC\_CHICK  
42 120 8.0 1828 2 Q700S8  
43 120 8.0 1828 2 CAF32691  
44 119.5 7.9 655 2 Q6CXQ4  
45 119.5 7.9 973 2 Q86TB8  
46 119.5 7.9 2042 2 Q6BUQ9  
47 119 7.9 895 2 Q9LIW7  
48 119 7.9 2664 2 Q7REL0  
49 118.5 7.9 335 2 Q7P7E4  
50 118 7.8 495 2 Q895W1

Q7shz4 neurospora  
Q8i754 plasmodium  
Q9szk7 arabidopsis  
Q7rfq7 plasmodium  
Q7raw8 plasmodium  
Q8n592 homo sapien  
Q60763 homo sapien  
Q7rd43 plasmodium  
Q6gne7 xenopus lae  
P29616 gallus gall  
Q700s8 mycoplasma  
Caf32691 mycoplasma  
Q6cxq4 kluyveromyc  
Q86tb8 homo sapien  
Q6bug9 debaryomyce  
Q9liw7 oryza sativ  
Q7rel0 plasmodium  
Q7p7e4 fusobacteri  
Q895w1 clostridium



```

FT MUTAGEN 182 185 Missing: In PMWK16; retained in cytosol.
FT MUTAGEN 182 183 Loss of function.
FT MUTAGEN 182 183 AG-DD: Abolishes cytotoxic activity.
FT MUTAGEN 186 187 AG-DD: Abolishes cytotoxic activity.
FT MUTAGEN 260 260 R->K: Strongly reduces cytotoxic activity.
FT MUTAGEN 267 267 D->A: Strongly reduces cytotoxic activity.

Query Match 99.7%; Score 1503; DB 1; Length 302;
Best Local Similarity 100.0%; Pred. No. 5.2e-87;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TEIVADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFQKQESQAASVL 61
Db 1 TEIVADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFQKQESQAASVL 60

Qy 62 VGDITKLMDSDQKYFEATQTVYEWCGVATOLLAAAYILLFDEYNEKASAKQDILIKVLD 121
Db 61 VGDITKLMDSDQKYFEATQTVYEWCGVATOLLAAAYILLFDEYNEKASAKQDILIKVLD 120

Qy 122 DGITKLNKAEQSLVSSQSFNNASGKLLDLSQTLNDFSEKSSYFQSQVDKIRKEAYAGA 181
Db 121 DGITKLNKAEQSLVSSQSFNNASGKLLDLSQTLNDFSEKSSYFQSQVDKIRKEAYAGA 180

Qy 182 AAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLSKVQNFPTTLNVTVKQANKDIDAAL 241
Db 181 AAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLSKVQNFPTTLNVTVKQANKDIDAAL 240

Qy 242 KLTTETAAIGEIKTETETTFYVDYDMLSLKKEAAKMNINTCNEYQKRGKKTLPFVP 301
Db 241 KLTTETAAIGEIKTETETTFYVDYDMLSLKKEAAKMNINTCNEYQKRGKKTLPFVP 300

Qy 302 EV 303
Db 301 EV 302

RESULT 2
HLYE ECO57
ID HLYE ECO57 STANDARD; PRT; 302 AA.
AC Q9REB3;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).
GN Name=hlyE; Synonyms=clyA, sheA; OrderedLocusNames=z1944, ECs1677;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EHEC;
RC MEDLINE=20322319; PubMed=10865950; DOI=10.1016/S0923-2508(00)00143-1;
RX del Castillo F.J., Moreno P., del Castillo I.;
RA "Characterization of the genes encoding the SheA haemolysin in
RT Escherichia coli O157:H7 and Shigella flexneri 2a.";
RL Res. Microbiol. 151:229-230(2000).
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RC MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RX Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RMD 0509952 / EHEC;
RC

```

```

RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RL O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -I- FUNCTION: Toxin, which has some hemolytic activity towards
CC mammalian cells. Acts by forming a pore-like structure upon
CC contact with mammalian cells (By similarity).
CC -I- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a
CC monomer, while in outer membrane vesicles, it oligomerizes to form
CC a pore structure that is active. Probably forms an octamer (By
CC similarity).
CC -I- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer
CC membrane vesicles. Also found in the periplasmic space (By
CC similarity).
CC -I- PTM: In periplasm, it forms a disulfide bond, which prevents the
CC oligomerization. In outer membrane vesicles, the redox status
CC prevents formation of the disulfide bond, leading to
CC oligomerization and pore formation (By similarity).
CC -I- SIMILARITY: Belongs to the hemolysin E family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ238954; CAB64962.1; ALT_INIT.
CC EMBL; AE005335; AAG56033.1; ALT_INIT.
CC EMBL; AP002555; BAB35100.1; ALT_INIT.
CC PIR; E85696; E85696.
CC PIR; E90838; E90838.
CC HSP; P77335; IQOY.
CC InterPro; IPR010356; HlyE.
CC Pfam; PF06109; HlyE; 1.
CC Complete proteome; Cytolysis; Hemolysis; Toxin; Transmembrane.
CC INIT MET 0 By similarity.
CC TRANSMEM 182 202 Potential
CC DISULFID 86 284 In monomeric form (By similarity).
CC SSEQUENCE 302 AA; 33585 MW; F261E29E1DE5FC87 CRC64;
CC
CC Query Match 98.7%; Score 1488; DB 1; Length 302;
CC Best Local Similarity 99.0%; Pred. No. 4.6e-86;
CC Matches 299; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
CC
CC Qy 2 TEIVADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFQKQESQAASVL 61
CC Db 1 TEIVADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFQKQESQAASVL 60
CC
CC Qy 62 VGDITKLMDSDQKYFEATQTVYEWCGVATOLLAAAYILLFDEYNEKASAKQDILIKVLD 121
CC Db 61 VGNIKTLMDSDQKYFEATQTVYEWCGVATOLLAAAYILLFDEYNEKASAKQDILIKVLD 120
CC
CC Qy 122 DGITKLNKAEQSLVSSQSFNNASGKLLDLSQTLNDFSEKSSYFQSQVDKIRKEAYAGA 181
CC Db 121 DGITKLNKAEQSLVSSQSFNNASGKLLDLSQTLNDFSEKSSYFQSQVDKIRKEAYAGA 180
CC
CC Qy 182 AAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLSKVQNFPTTLNVTVKQANKDIDAAL 241
CC Db 181 AAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLSKVQNFPTTLNVTVKQANKDIDAAL 240
CC
CC Qy 242 KLTTETAAIGEIKTETETTFYVDYDMLSLKKEAAKMNINTCNEYQKRGKKTLPFVP 301
CC Db 241 KLTTETAAIGEIKTETETTFYVDYDMLSLKKEAAKMNINTCNEYQKRGKKTLPFVP 300
CC
CC Qy 302 EV 303
CC Db 301 EV 302

```

DR EMBL; AJ313034; CAC38363.1; --  
 DR EMBL; AL627270; CAD01758.1; ALT\_INIT.  
 DR EMBL; AE016839; AAO69115.1; ALT\_INIT.  
 DR HSP; P77335; 100Y.  
 DR InterPro; IPR010356; HlyE.  
 DR Pfam; PF06109; HlyE; 1.  
 KW Complete proteome; Cytolysis; Hemolysis; Toxin; Transmembrane.  
 FT INIT MET 0 By similarity.  
 FT TRANSMEM 178 198 Potential.  
 FT DISULFID 86 284 In monomeric form (By similarity).  
 SQ SEQUENCE 302 AA; 33658 MW; 29DFD835D04FDAE CRC64;

Query Match 92.5%; Score 1395; DB 1; Length 302;  
 Best Local Similarity 91.1%; Pred. NO. 3.4e-80;  
 Matches 275; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

QY 2 TEIVADKTVEVVKNAETADGALDLYNKYLQDVIPWQTFDETIKELSRFKQEYSQAASVL 61  
 DB 1 TGIFAEQTEVVVKSAETADGALDLYNKYLQDVIPWQTFDETIKELSRFKQEYSQAASVL 60  
 QY 62 VGDIKTLMDSDQKYFEATQTVYVCGVATQLLAAAYILLFDEYNEKKASAKDILITVLD 121  
 DB 61 VGDIKVLLMDSDQKYFEATQTVYVCGVATQLLAAAYILLFDEYNEKKASAKDILITVLD 120  
 QY 122 DGIITKLENAQKSLIVSSQSFNNASGKLIALDSQTNDFSEKSSYFQSDVKIRKEAYAGA 181  
 DB 121 DGVKKLENAQKSLIVSSQSFNNASGKLIALDSQTNDFSEKSSYFQSDVKIRKEAYAGA 180  
 QY 182 AGAVVAGPGLIISYSIAAGVVEGKLIPELKNKLSVQNFVTTLSNTVKQANKDIDAAL 241  
 DB 181 AGIVVAGPGLIISYSIAAGVVEGKLIPELKNKLSVQNFVTTLSNTVKQANKDIDAAL 240  
 QY 242 KLTTEIAAIGEITETETTRFVYVDVDDMLSLKKAQKQNTCNQYKSHGKKTLPFVP 301  
 DB 241 KLATEIAAIGEITETETTRFVYVDVDDMLSLKKAQKQNTCNQYKSHGKKTLPFVP 300  
 QY 302 EV 303  
 DB 301 DV 302

RESULT 4  
 HLYE\_SALPA STANDARD; PRT; 302 AA.  
 ID HLYE\_SALPA  
 AC Q9IRRG; 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).  
 GN Name=hlyE; Synonyms=clyA, sheA;  
 OS Salmonella paratyphi-a  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=54388;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SMI S2068/99;  
 RX MEDLINE=22215712; PubMed=12228306;  
 RA Oscarsson J., Westmark M., Loeftdahl S., Olsen B., Palmgren H.,  
 RA Mizunoe Y., Wai S.N., Uhlin B.E.;  
 RT "Characterization of a pore-forming cytotoxin expressed by Salmonella  
 RL Infect. Immun. 70:5759-5769(2002).  
 CC -!- FUNCTION: Toxin, which has some hemolytic activity towards  
 CC mammalian cells. Acts by forming a pore-like structure upon  
 CC contact with mammalian cells (By similarity).  
 CC -!- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a  
 CC monomer, while in outer membrane vesicles, it oligomerizes to form  
 CC a pore structure that is active. Probably forms an octamer (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer  
 CC membrane vesicles. Also found in the periplasmic space (By  
 CC similarity).  
 CC EMBL; AJ313032; CAC38360.1; --

RESULT 3  
 HLYE\_SALTI STANDARD; PRT; 302 AA.  
 ID HLYE\_SALTI  
 AC Q82727; Q934C4; 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).  
 GN Name=hlyE; Synonyms=clyA, sheA; OrderedLocusNames=STY1498, t1477;  
 OS Salmonella typhi  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ty21a and SMI S2369/96;  
 RX MEDLINE=22215712; PubMed=12228306;  
 RA Oscarsson J., Westmark M., Loeftdahl S., Olsen B., Palmgren H.,  
 RA Mizunoe Y., Wai S.N., Uhlin B.E.;  
 RT "Characterization of a pore-forming cytotoxin expressed by Salmonella  
 RL Infect. Immun. 70:5759-5769(2002).  
 CC -!- FUNCTION: Toxin, which has some hemolytic activity towards  
 CC mammalian cells. Acts by forming a pore-like structure upon  
 CC contact with mammalian cells (By similarity).  
 CC -!- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a  
 CC monomer, while in outer membrane vesicles, it oligomerizes to form  
 CC a pore structure that is active. Probably forms an octamer (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer  
 CC membrane vesicles. Also found in the periplasmic space (By  
 CC similarity).  
 CC -!- PTM: In periplasm, it forms a disulfide bond, which prevents the  
 CC oligomerization. In outer membrane vesicles, the redox status  
 CC prevents formation of the disulfide bond, leading to  
 CC oligomerization and pore formation (By similarity).  
 CC -!- SIMILARITY: Belongs to the hemolysin E family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AJ313032; CAC38360.1; --



CC -|- PTM: In periplasm, it forms a disulfide bond, which prevents the  
 CC oligomerization. In outer membrane vesicles, the redox status  
 CC prevents formation of the disulfide bond, leading to  
 CC oligomerization and pore formation (By similarity).  
 CC -|- SIMILARITY: Belongs to the hemolysin E family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AJ313033; CAC38362.1; -  
 DR HSP; P77335; 100Y.  
 DR InterPro; IPR010356; HlyE.  
 DR Pfam; PF06109; HlyE; 1.  
 KW Cytolysis; Hemolysis; Toxin; Transmembrane.  
 FT INIT\_MET 0  
 FT TRANSMEM 178 198 By similarity.  
 FT DISULFID 86 284 Potential.  
 FT SEQUENCE 302 AA; 33613 MW; 22DB9112B763A061 CRC64;  
 SQ  
 Query Match 91.28; Score 1376; DB 1; Length 302;  
 Best Local Similarity 89.78; Pred. No. 5.4e-79;  
 Matches 271; Conservative 18; Mismatches 13; Indels 0; Gaps 0;  
 QY 2 TEIVADKTVVKNNAIETADGALDLYNKYLDQVTPWOTFDTIKELSRFKQEYSQAASVL 61  
 DB 1 TGIFAEQTVVKNNAIETADGALDLYNKYLDQVTPWOTFDTIKELSRFKQEYSQAASVL 60  
 QY 62 VGDIKTLMDSDQKYFEATQTVYVWCGVATQLLAAYILLFDEYNEKKASAKQDILIKVLD 121  
 DB 61 VGDIKTLMDSDQKYFEATQTVYVWCGVATQLLAAYILLFDEYNEKKASAKQDILIKVLD 120  
 QY 122 DGIITKLEAKQSLVSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDKIRKEAYAGA 181  
 DB 121 DGVNKLKLEAKQSLVSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDKIRKEAYAGA 180  
 QY 182 AAGVAVPFGLLIISYSIAAGVVEGKLIPELKNKLSQNFPTTILSNVTKQAKDIDAACL 241  
 DB 181 AAGVAVPFGLLIISYSIAAGVVEGKLIPELKNKLSQNFPTTILSNVTKQAKDIDAACL 240  
 QY 242 KLTETIAAIGEIKTETETTRFFVYDDMLSLKKAARKMINTCNEYQKRGHKKTFEVP 301  
 DB 241 KLTETIAAIGEIKTETETTRFFVYDDMLSLKKAARKMINTCNEYQKRGHKKTFEVP 300  
 QY 302 EV 303  
 DB 301 DI 302  
 RESULT 5  
 Q9X2S8 PRELIMINARY; PRT; 300 AA.  
 AC Q9X2S8;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hemolysin.  
 GN Name=HlyE;  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=964660;  
 RX MEDLINE=99242013; PubMed=10227474;  
 RA Reingold J., Starr N., Maurer J., Lee M.D.;  
 RT "Identification of a new Escherichia coli She haemolysin homolog in  
 RT avian E. coli.";

RL Vet. Microbiol. 66:125-134(1999).  
 DR EMBL; AF052225; AAD28079.1; -  
 DR HSP; P77335; 100Y.  
 DR InterPro; IPR010356; HlyE.  
 DR Pfam; PF06109; HlyE; 1.  
 SQ SEQUENCE 300 AA; 33555 MW; 6580B66C4A7B4BC CRC64;  
 Query Match 75.18; Score 1133; DB 2; Length 300;  
 Best Local Similarity 76.98; Pred. No. 1.2e-63;  
 Matches 223; Conservative 31; Mismatches 36; Indels 0; Gaps 0;  
 QY 6 ADKTVEVKNNAIETADGALDLYNKYLDQVTPWOTFDTIKELSRFKQEYSQAASVLVGD 65  
 DB 4 ADQTVETVKNNAIETADGALDLYNKYLDQVTPWOTFDTIKELSRFKQEYSQAASVLVGEI 63  
 QY 66 KTLMDSDQKYFEATQTVYVWCGVATQLLAAYILLFDEYNEKKASAKQDILIKVLDG 125  
 DB 64 KSLMNSQDRYPEATQTVYVWCGVATQLLAAYILLFDEYNEKKASAKQDILIKVLDG 123  
 QY 126 KLEAKQSLVSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDKIRKEAYAGA 185  
 DB 124 KLEAKQSLVSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDKIRKEAYAGA 183  
 QY 186 VAGPFGLLIISYSIAAGVVEGKLIPELKNKLSQNFPTTILSNVTKQAKDIDAACL 245  
 DB 184 VGRPFGLLIISYSIAAGVVEGKLIPELKNKLSQNFPTTILSNVTKQAKDIDAACL 243  
 QY 246 EIAAIGEIKTETETTRFFVYDDMLSLKKAARKMINTCNEYQKRGHKK 295  
 DB 244 EISVIGDKLTETETTRFFVYDDMLSLKKAARKMINTCNEYQKRGHKK 293  
 RESULT 6  
 HLEL SHIFL  
 ID HLEL SHIFL STANDARD; PRT; 113 AA.  
 AC Q9RGT3; Q9RGT3;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Hemolysin E-like protein.  
 GN OrderedLocustNames=SF1171, S1259;  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Shigella.  
 OX NCBI\_TaxID=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12022 / Serotype 2b;  
 RX MEDLINE=20123445; PubMed=10660049;  
 RA Wallace A.J., Stillman T.J., Atkins A., Jamieson S.J., Bullough P.A.,  
 RA Green J., Artymuk P.J.;  
 RT "E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of  
 RT the toxin and observation of membrane pores by electron microscopy.";  
 RL Cell 100:265-276(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CECT 585 / Serotype 2a;  
 RX MEDLINE=20322319; PubMed=10865950; DOI=10.1016/S0923-2508(00)00143-1;  
 RA del Castillo F.J., Moreno F., del Castillo I.;  
 RT "Characterization of the genes encoding the SheA haemolysin in  
 RT Escherichia coli O157:H7 and Shigella flexneri 2a.";  
 RL Res. Microbiol. 151:229-230(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=301 / Serotype 2a;  
 RX MEDLINE=22272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu K., Kan B., Ding K., Hou Y.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Chen Y.,  
 RA Yu J.;  
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
 RT through comparison with genomes of Escherichia coli K12 and O157.";

Nucleic Acids Res. 30:4432-4441(2002).

[4]  
 SEQUENCE FROM N.A.  
 STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 MEDLINE=22590274; Pubmed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RT "Complete genome sequence and comparative genomics of Shigella  
 RT flexneri serotype 2a strain 2457T";  
 RL Infect. Immun. 71:2775-2786(2003).  
 CC -1- SIMILARITY: Belongs to the hemolysin E family.  
 CC -1- CAUTION: Although it is strongly related to the hemolysin E toxin  
 CC from E.coli K-12 strain, it lacks all the C-terminal part of the  
 CC protein, due to a deletion that creates a frameshift, and it is  
 CC therefore not functional. May be a pseudogene.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF200955; AAF13995.1; -  
 CC EMBL; AJ238955; CAB65415.1; ALT INIT.  
 CC EMBL; AE015143; -; NOT ANNOTATED\_CDS.  
 CC EMBL; AE016982; -; NOT ANNOTATED\_CDS.  
 CC HSP; P77335; 100Y.  
 CC InterPro; IPR010356; HlyE.  
 CC Pfam; PF06109; HlyE; 1.  
 CC SEQUENCE 113 AA; 12879 MW; C328908D14C5C4EB CRC64;  
 Query Match 37.3%; Score 563; DB 1; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-28;  
 Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 1 MTEIVADKTVVEVKNAIETADGALDLYNKYLDQVWPQDETTELKSRFKQYSQASV 60  
 D 1 MTEIVADKTVVEVKNAIETADGALDLYNKYLDQVWPQDETTELKSRFKQYSQASV 60  
 Y 61 LVGDIKLLMDSQDKYFEATQVYEWGCVATQLLAAVILLFDEYNEKKASA 111  
 D 61 LVGDIKLLMDSQDKYFEATQVYEWGCVATQLLAAVILLFDEYNEKKASA 111  
 RESULT 7  
 HLEL ECOL6  
 ID HLEL ECOL6 STANDARD; PRT; 93 AA.  
 AC Q8FI27;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DE Hemolysin E-like protein.  
 GN OrderedLocusNames=cl630;  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=O6:H1 / CFT073 / ATCC 700928 / UPEC;  
 RX MEDLINE=2238234; Pubmed=12471157; DOI=10.1073/pnas.252529799;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 CC -1- SIMILARITY: Belongs to the hemolysin E family.

CC -1- CAUTION: Although it is strongly related to the hemolysin E toxin  
 CC from E.coli K-12 strain, it lacks all the N-terminal part of the  
 CC protein, and it is therefore probably not functional. May be a  
 CC pseudogene.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE016759; AAN80095.1; -  
 CC HSP; P77335; 100Y.  
 CC InterPro; IPR010356; HlyE.  
 CC Pfam; PF06109; HlyE; 1.  
 CC Complete proteome.  
 CC SEQUENCE 93 AA; 10590 MW; CB1625028DC36DE6 CRC64;  
 Query Match 29.0%; Score 437; DB 1; Length 93;  
 Best Local Similarity 92.5%; Pred. No. 2.5e-20;  
 Matches 86; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 Y 211 LKMKLSVONFPTLNTVQKANDDAKLLTETIAAIGETETTRFRFYVDVDDL 270  
 D 1 MKMKLSALSFTLNTVQKANDDAKLLTETIAAIGETETTRFRFYVDVDDL 60  
 Y 271 LSLLEKAAKQKMTNCNEYKRGKTKLFEVPEV 303  
 D 61 LSLLEKAAKQKMTNCNEYKRGKTKLFEVPEV 93  
 RESULT 8  
 SYR SYNY3  
 ID SYR SYNY3 STANDARD; PRT; 584 AA.  
 AC Q55486;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).  
 GN Name=ARGS; OrderedLocusNames=sll10502;  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96127529; Pubmed=8590279;  
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
 RA Sugita M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 64% to 92% of the genome.";  
 RL DNA Res. 2:153-166(1995).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +  
 CC diphosphate + L-arginyl-tRNA(Arg).  
 CC -1- SUBUNIT: Monomer (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D64006; BAA10833.1; -  
 CC PIR; S75986; S75986.  
 CC HSP; Q93RP5; 11Q0.  
 CC HAMAP; MF 00123; -; 1.  
 CC InterPro; IPR005148; ArgtrNASynthet\_N.



|  |  |   |     |
|--|--|---|-----|
| QY   | 115  | ILIKVLDDGITKLNBAOK-----SLLVSSQSFNNAAGSKLLALDSQLTNDPSE       | 161 |
| DB   | 342  | DLNKKREDLINRIASSELAKKABELNTKLVLFVKQBALNKKSGQYLYYINKLNDLELR  | 401 |
| QY   | 162  | KSSYFQSQVDKIRKEAYAGAAAGVAGVPGFGLIISVSIAGVVEGKLIPELKNKLSQVNF | 221 |
| DB   | 402  | LADKYKNSNKLSR-----LKNHIGEYNKQ                               | 426 |
| QY   | 222  | FTTSLNTVQANKIDAAKLTTEIAAIGETETTRFVYDDYDDIMLMLSLLEKAAKGM     | 281 |
| DB   | 427  | LEKIENELBECNKIDNTKKQL-AEFDKSNKKQOELE-----SELVQLNKKI         | 472 |
| QY   | 282  | INTCNVYQKRH-----GKTLFEVPEV                                  | 303 |
| DB   | 473  | ----DELGRKHKRQLEASOKKALDEAKI                                | 499 |
| RESULT 12  |  |   |     |
| ID   | Q9X360   | PRELIMINARY; PRT; 652 AA.                                   |     |
| AC   | Q9X360;  |   |     |
| DT   | 01-NOV-1999  | (TremBLrel. 12, Created)                                    |     |
| DT   | 01-NOV-1999  | (TremBLrel. 12, Last sequence update)                       |     |
| DT   | 01-OCT-2004  | (TremBLrel. 28, Last annotation update)                     |     |
| DE   | PX01-90  | (S-layer protein,).   |     |
| GN   | OrderedLocusNames=GBAA_PX01_0124;                                      |   |     |
| OS   | Bacillus anthracis.  |   |     |
| OG   | Plasmid virulence plasmid pX01, and Plasmid pX01.                      |   |     |
| OC   | Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.               |   |     |
| OX   | NCBI_TaxID=1392;   |   |     |
| RN   | [1]  |   |     |
| RP   | SEQUENCE FROM N.A.   |   |     |
| RC   | STRAIN=Sterner; PLASMID=virulence plasmid PX01;                        |   |     |
| RX   | MEDLINE=99445483; PubMed=10515943;                                     |   |     |
| RA   | Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,        |   |     |
| RA   | Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Manter D.,    |   |     |
| RA   | Martinez Y., Ricke D., Svensson R., Jackson P.J.;                      |   |     |
| RT   | "Sequence and organization of pX01, the large Bacillus anthracis       |   |     |
| RT   | plasmid harboring the anthrax toxin genes.";                           |   |     |
| RL   | J. Bacteriol. 181:6509-6515(1999).                                     |   |     |
| RN   | [2]  |   |     |
| RP   | SEQUENCE FROM N.A.   |   |     |
| RC   | STRAIN=Ames / isolate 0581; PLASMID=pX01;                              |   |     |
| RA   | Ravel J., Raeko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B., |   |     |
| RA   | Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,            |   |     |
| RA   | Fraser C.M.;   |   |     |
| RT   | "Bacillus anthracis comparative genomics.";                            |   |     |
| RL   | Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.                |   |     |
| DR   | EMBL; AF065404; AAD32394.1; -.   |   |     |
| DR   | EMBL; AB017336; AAT28865.2; -.   |   |     |
| DR   | PIR; B59102; B59102.   |   |     |
| DR   | TIGR; GBAA_PX01_0124; -.   |   |     |
| DR   | InterPro; IPR001119; SLH.  |   |     |
| DR   | Pfam; PF00395; SLH; 3.   |   |     |
| DR   | PROSITE; PS01072; SLH_DOMAIN; UNKNOWN_1.                               |   |     |
| KW   | Complete proteome; Plasmid.  |   |     |
| SQ   | SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;                     |   |     |
| Query Match 8.4%; Score 126; DB 2; Length 652;                   |  |   |     |
| Best Local Similarity 20.5%; Pred. No. 10;                       |  |   |     |
| Matches 68; Conservative 57; Mismatches 108; Indels 98; Gaps 11; |  |   |     |
| QY   | 7  | DKTVEVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQVYSQAASVLVGDIK | 66  |
| DB   | 233  | NKRRDINQKIDFED-KLNSQRDLERML-----EELNKLSQLKQSQPQ-----LQDLK   | 281 |
| QY   | 67   | TLLMSQDKYFEATQ-----TVYEWGCVATQLLAAAYILLDFEYNE--KKASAKD      | 114 |
| DB   | 282  | NKLKESQSRLELLELNKDSNRLEINSEIKLNDRAKELLIMELIKQOSEPDKIKNEKD   | 341 |
| QY   | 115  | ILIKVLDDGITKLNBAOK-----SLLVSSQSFNNAAGSKLLALDSQLTNDPSE       | 161 |
| DB   | 342  | DLNKKREDLINRIASSELAKKABELNTKLVLFVKQBALNKKSGQYLYYINKLNDLELR  | 401 |

|  |  |  |     |
|--|--|--|-----|
| QY   | 2  | TEIVADTKVEVVKNAIETADGALDLYNKYLDQVWPQTF-----DETIKELSR         | 49  |
| DB   | 178  | TGIIVDLKLISSKERLNLNLELLKKYVKGQDQAIKAVTSAIMRSRSGIKNPKDPGIFSLF | 237 |
| QY   | 50   | F-----KQVYSQA-ASVLVGDITKLLMSQDKYFEATQTV-----YEWGCVAT         | 91  |
| DB   | 238  | FGPTGVGKTEVARSLADILFNSPKKMIRLDMSEYMERKSHVAKLIGAPPGVGVGEEGRLT | 297 |
| QY   | 92   | QLLAA---YILLPEYNEKASAKDILIKVLDDG-----ITKLNBAOKSLLVSSQSFFN    | 143 |
| DB   | 298  | EAVARNPYSILFDEI-EKAHSDVFNILLQLDDGRLTDSLGKTIYDFKNTIIWVTS--NI  | 354 |
| QY   | 144  | ASGKLLALDSQLTNDPSE-KSSYFQSQVDKIRKEAYAGAAAGVAGVPGFGLIISVSIAGV | 202 |
| DB   | 355  | ASQYLLTSD-----DFVQIDDKIQAEINQTFRPEFLNRIDNIVY--FNALSVQTI--GE  | 405 |
| QY   | 203  | VEGKLIPELKNKLSQVNFPTLNTVQKANKIDAAKLTTEIAAIGETETTRF           | 262 |
| DB   | 406  | IVDKLLDELITRLQDEQNYFINFS-----EARNKIINE-----                  | 439 |
| QY   | 263  | YVDYDDLML-LSLLKEAAKMINT                                      | 284 |
| DB   | 440  | --GYDRLFGARLLKDLKLIET  | 460 |
| RESULT 11  |  |  |     |
| ID   | Q7CMF0   | PRELIMINARY; PRT; 652 AA.                                    |     |
| AC   | Q7CMF0;  |  |     |
| DT   | 05-JUL-2004  | (TremBLrel. 27, Created)                                     |     |
| DT   | 05-JUL-2004  | (TremBLrel. 27, Last sequence update)                        |     |
| DT   | 05-JUL-2004  | (TremBLrel. 27, Last annotation update)                      |     |
| DE   | S-layer protein, (PX01-90).  |  |     |
| GN   | Name=BXA0124;  |  |     |
| OS   | Bacillus anthracis str. A2012.   |  |     |
| OG   | Plasmid pX01.  |  |     |
| OC   | Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.               |  |     |
| OX   | NCBI_TaxID=191218;   |  |     |
| RN   | [1]  |  |     |
| RP   | SEQUENCE FROM N.A.   |  |     |
| RC   | STRAIN=A2012;  |  |     |
| RX   | MEDLINE=22061436; PubMed=12004073;                                     |  |     |
| RA   | Read T.D., Salzberg S.L., Pop M., Shumway M., Umayam L., Jiang L.,     |  |     |
| RA   | Holtzapple E., Busch J.D., Smith K.L., Schupp J.M., Solomon D.,        |  |     |
| RA   | Keim P., Fraser C.M.;  |  |     |
| RT   | "Comparative Genome Sequencing for Discovery of Novel Polymorphisms in |  |     |
| RL   | Bacillus anthracis."   |  |     |
| DR   | Science 296:2028-2033(2002).   |  |     |
| DR   | EMBL; AF011190; AAM26077.1; -.   |  |     |
| DR   | InterPro; IPR001119; SLH.  |  |     |
| DR   | InterPro; IPR010989; t-snare.  |  |     |
| DR   | InterPro; IPR010978; tRNA_binding_arm.                                 |  |     |
| DR   | Pfam; PF00395; SLH; 3.   |  |     |
| DR   | PROSITE; PS01072; SLH_DOMAIN; UNKNOWN_1.                               |  |     |
| KW   | Plasmid.   |  |     |
| SQ   | SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;                     |  |     |
| Query Match 8.4%; Score 126; DB 2; Length 652;                   |  |  |     |
| Best Local Similarity 20.5%; Pred. No. 10;                       |  |  |     |
| Matches 68; Conservative 57; Mismatches 108; Indels 98; Gaps 11; |  |  |     |
| QY   | 7  | DKTVEVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQVYSQAASVLVGDIK  | 66  |
| DB   | 233  | NKRRDINQKIDFED-KLNSQRDLERML-----EELNKLSQLKQSQPQ-----LQDLK    | 281 |
| QY   | 67   | TLLMSQDKYFEATQ-----TVYEWGCVATQLLAAAYILLDFEYNE--KKASAKD       | 114 |
| DB   | 282  | NKLKESQSRLELLELNKDSNRLEINSEIKLNDRAKELLIMELIKQOSEPDKIKNEKD    | 341 |

```
QY 162 KSSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYTAAGVVEGKLIPELKNKLKSVQNF 221
D 402 LADKYNSDNKISR-----LKNHIGYENKQ 426
QY 222 FTTLSTNTVKANKIDAAKLLKLTETIAAIGBIKTETTRFYVDYDDMLSLKKEAAKQM 281
D 427 LEKIENELECNKIDNTKKQL-AEFDKSNKKQOELE-----SELVOLNKKI 472
QY 282 INTNEYQKRH-----GKTKLFVPEV 303
D 473 ----DELGKGRKHQOELEASQKKALDEAKEI 499
RESULT 13
AAT28865 PRELIMINARY; PRT; 652 AA.
AC AAT28865;
DT 01-JUN-2004 (TrEMBLrel. 27, Created)
DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE S-layer protein, (PXOI-90).
GN GBAA_PXOI_0124.
OS Bacillus anthracis str. Ames 0581.
OG Plasmid pXOI.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames 0581.
RA Ravel J., Raeko D.A., Shumway M.P., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE017336; AAT28865.2; -.
KW Plasmid
SQ SEQUENCE 652 AA; 76210 MW; 723F5FB03516355 CRC64;
Query Match 8.4%; Score 126; DB 2; Length 652;
Best Local Similarity 20.5%; Pred. No. 10;
Matches 68; Conservative 57; Mismatches 108; Indels 98; Gaps 11;
QY 7 DKTVEVVKNAIETADGALDLYNKYLDQVIFWQTFDETIKELSRFKQYSQAASVLVDGIK 66
D 233 NKERDVINQIDBFD-KLNSQRKDLERML-----BELNQLSQLKQSQPQ-----LQDLK 281
QY 67 TLLMDSQDKYPEATO-----TVYWCVGATQLLAAYVILLFDEYNE--KKASAQKD 114
D 282 NLKESQSRLELNKDSNLELNSEIKLNDKRAELLSLIMELIKQSEFDRKIRNEKD 341
QY 115 ILIKVLDDGITKLNEAQK-----SLIVSSQSFNNASGKLLALDSQLTNDPSE 161
D 342 DLNKKREDLNRIAESKELAKKKAELNTKLVELFKVQELNKKSGQVLYYINKLDNELRE 401
QY 162 KSSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYTAAGVVEGKLIPELKNKLKSVQNF 221
D 402 LADKYNSDNKISR-----LKNHIGYENKQ 426
QY 222 FTTLSTNTVKANKIDAAKLLKLTETIAAIGBIKTETTRFYVDYDDMLSLKKEAAKQM 281
D 427 LEKIENELECNKIDNTKKQL-AEFDKSNKKQOELE-----SELVOLNKKI 472
QY 282 INTNEYQKRH-----GKTKLFVPEV 303
D 473 ----DELGKGRKHQOELEASQKKALDEAKEI 499
RESULT 14
Q9EP71 PRELIMINARY; PRT; 979 AA.
AC Q9EP71;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
```

```
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE NORPEG-like protein (Ankycorbin).
GN Name=Rail4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21264742; PubMed=11042181;
RA Kuty R.K., Kuty G., Samuel W., Duncan T., Bridges C.C.,
RA El-Sherbeeny A., Naginei C.N., Smith S.B., Wiggert B.;
RT "Molecular characterization and developmental expression of NORPEG, a
RT novel gene induced by retinoic acid.";
RL J. Biol. Chem. 276:2831-2840(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Maman A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57 Black/6;
RX MEDLINE=21110708; PubMed=11168586;
RA Peng Y.F., Mandai K., Sakisaka T., Okabe N., Yamamoto Y., Yokoyama S.,
RA Mizoguchi A., Shiozaki H., Monden M., Takai Y.;
RT "Ankycorbin: a novel actin cytoskeleton-associated protein.";
RL Genes Cells 5:1001-1008(2000).
DR EMBL: AF274866; AAG25937.1; -.
DR EMBL: BC052458; AAG2458.1; -.
DR EMBL: AF202315; AAG24483.1; -.
DR HSSP: P42773; 1IHB.
DR MGD: MGI:1922896; Rail4.
DR GO: GO:0005739; C:mitochondrion; IDA.
DR InterPro: IPR002110; ANK.
DR Pfam: PF00023; Ank; 7.
DR PRINTS: PR01415; ANKYRIN.
DR SMART: SM00248; ANK; 6.
DR PROSITE: PS50088; ANK_REPEAT; 5.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
KW ANK repeat.
SQ SEQUENCE 979 AA; 108851 MW; B2E8C016D80237C4 CRC64;
Query Match 8.4%; Score 126; DB 2; Length 979;
Best Local Similarity 21.5%; Pred. No. 17;
Matches 78; Conservative 64; Mismatches 125; Indels 96; Gaps 17;
```

QY 1 MTEIVA-DKTVEVVKNAIETADGALDLYNKYLDQVIFW--QTDETIKELSRFKQYSQA 57

```

Db      497 MKEVLSVQKQKGLLSQESADGYSHLREAPDEDIDTLKQDLQKAVESARKER-----S 552
QY      58 ASVLVGDIKTLMDSDQKYEATQTVYEWGCVATQLAAAYILLFDEYNKKA-----S 110
Db      553 -----VRELETKLAKEQA--EATKPPAEAC---EELRSSYCSVIENNKEKAPLFEKYQQ 603
QY      111 AQKDILIKVLDDGKITKLENAQKSLV-----SSQSFNNASGKLLALDLSQTLNDFSEKSSY 165
Db      604 AQEIM-----KLKDTLKSQMPQEPAPDSDGMKEAMNRM-----DELNKKVSELSQL 651
QY      166 F---OSQVDKIRK-EAYAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLKSVON- 220
Db      652 YREAQAELEDYRKRSLSLEDAAYIHKAEHERLMHVSNLSRAKSEALSEMKSQYSKVLNE 711
QY      221 -----FFTLTSLNTVKQ-----ANKDIDAACL--KL 243
Db      712 LTQLKQLVDAHKENSVSITEHLQVITTLRTAKEMEKEKISALTCHLANKEAEVAKLEKQL 771
QY      244 TTEIAAIGEIKTETTRFYVDYDDLMLSLKEA---AKMINTCNEYQKHGKKTILFEV 300
Db      772 AEEKAASVDMVPKSS-----YEKQASLESEVNALATKLKESVREKRAHS-----EV 820
QY      301 PEV 303
Db      821 AQV 823

RESULT 15
Q6ZPT6 PRELIMINARY; PRT; 992 AA.
ID Q6ZPT6 AC Q6ZPT6;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DE MKIAA1334 protein (Fragment).
GN Name=MKIAA1334;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RC TISSUE=Embryonic tail;
RX PubMed=14621295;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
Saga Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT III. The complete nucleotide sequences of 500 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries."
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129333; BAC98143.1;
DR GO; GO:0005739; C-mitochondrion; IDA.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; Ank; 7.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SMO0248; ANK; 6.
DR PROSITE; PS50088; ANK_REPEAT; 5.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
KW ANK repeat.
SQ SEQUENCE 992 AA; 110311 MW; FBBC2380E775EB5B CRC64;
Best Local Similarity 8.4%; Score 126; DB 2; Length 992;
Matches 78; Conservative 64; Mismatches 125; Indels 96; Gaps 17;

QY      1 MTEIVA-DKTVVVKNAIETADGALDLYNKYLQVTPW--QTDETIKLSRFEKQYSQA 57
Db      510 MKEVLSVQKQKGLLSQESADGYSHLREAPDEDIDTLKQDLQKAVESARKER-----S 565
QY      58 ASVLVGDIKTLMDSDQKYEATQTVYEWGCVATQLAAAYILLFDEYNKKA-----S 110
Db      566 -----VRELETKLAKEQA--EATKPPAEAC---EELRSSYCSVIENNKEKAPLFEKYQQ 616
QY      111 AQKDILIKVLDDGKITKLENAQKSLV-----SSQSFNNASGKLLALDLSQTLNDFSEKSSY 165
Db      617 AQEIM-----KLKDTLKSQMPQEPAPDSDGMKEAMNRM-----DELNKKVSELSQL 664
QY      166 F---OSQVDKIRK-EAYAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLKSVON- 220
Db      665 YREAQAELEDYRKRSLSLEDAAYIHKAEHERLMHVSNLSRAKSEALSEMKSQYSKVLNE 724
QY      221 -----FFTLTSLNTVKQ-----ANKDIDAACL--KL 243
Db      821 AQV 823

```

```

Db      566 -----VRELETKLAKEQA--EATKPPAEAC---EELRSSYCSVIENNKEKAPLFEKYQQ 616
QY      111 AQKDILIKVLDDGKITKLENAQKSLV-----SSQSFNNASGKLLALDLSQTLNDFSEKSSY 165
Db      617 AQEIM-----KLKDTLKSQMPQEPAPDSDGMKEAMNRM-----DELNKKVSELSQL 664
QY      166 F---OSQVDKIRK-EAYAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLKSVON- 220
Db      665 YREAQAELEDYRKRSLSLEDAAYIHKAEHERLMHVSNLSRAKSEALSEMKSQYSKVLNE 724
QY      221 -----FFTLTSLNTVKQ-----ANKDIDAACL--KL 243
Db      725 LTQLKQLVDAHKENSVSITEHLQVITTLRTAKEMEKEKISALTCHLANKEAEVAKLEKQL 784
QY      244 TTEIAAIGEIKTETTRFYVDYDDLMLSLKEA---AKMINTCNEYQKHGKKTILFEV 300
Db      785 AEEKAASVDMVPKSS-----YEKQASLESEVNALATKLKESVREKRAHS-----EV 833
QY      301 PEV 303
Db      834 AQV 836

RESULT 16
BAC98143 PRELIMINARY; PRT; 992 AA.
ID BAC98143 AC BAC98143;
DT 02-MAR-2004 (TREMELrel. 27, Created)
DT 02-MAR-2004 (TREMELrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMELrel. 27, Last annotation update)
DE MKIAA1334 protein (Fragment).
GN MKIAA1334.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RC TISSUE=Embryonic tail;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
Saga Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene:
RT III. The Complete Nucleotide Sequences of 500 Mouse KIAA-homologous
RT cDNAs Identified by Screening of Terminal Sequences of cDNA Clones
RT Randomly Sampled from Size-fractionated Libraries."
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129333; BAC98143.1;
DR NON TER 1
SQ SEQUENCE 992 AA; 110311 MW; FBBC2380E775EB5B CRC64;
Query Match 8.4%; Score 126; DB 2; Length 992;
Best Local Similarity 21.5%; Pred. No. 17;
Matches 78; Conservative 64; Mismatches 125; Indels 96; Gaps 17;

QY      1 MTEIVA-DKTVVVKNAIETADGALDLYNKYLQVTPW--QTDETIKLSRFEKQYSQA 57
Db      510 MKEVLSVQKQKGLLSQESADGYSHLREAPDEDIDTLKQDLQKAVESARKER-----S 565
QY      58 ASVLVGDIKTLMDSDQKYEATQTVYEWGCVATQLAAAYILLFDEYNKKA-----S 110
Db      566 -----VRELETKLAKEQA--EATKPPAEAC---EELRSSYCSVIENNKEKAPLFEKYQQ 616
QY      111 AQKDILIKVLDDGKITKLENAQKSLV-----SSQSFNNASGKLLALDLSQTLNDFSEKSSY 165
Db      617 AQEIM-----KLKDTLKSQMPQEPAPDSDGMKEAMNRM-----DELNKKVSELSQL 664
QY      166 F---OSQVDKIRK-EAYAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLKSVON- 220
Db      665 YREAQAELEDYRKRSLSLEDAAYIHKAEHERLMHVSNLSRAKSEALSEMKSQYSKVLNE 724
QY      221 -----FFTLTSLNTVKQ-----ANKDIDAACL--KL 243

```

```
Db 725 LTQLKQLVDAHKENSVSITEHLQVITTTTRITAKEMBEKISALTCHLANKEAEVAKLEKQL 784
Qy 244 TTEIAAGETIKETETTRFVDDVDDMLSLKKA---AKMINTCNEYQKRGHKTKLFEV 300
Db 785 ABEKAASVDAMPKSS-----YEKLOASLESEVNALATKLKESVREKRAHS-----EV 833
Qy 301 PEV 303
Db 834 AQV 836

RESULT 17
ID Q81752 PRELIMINARY; PRT; 504 AA.
AC Q81752;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Rhopty protein (Fragment).
GN Name-rhop235-15;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RA Khan S.M., Jarra W., Bayele H., Preiser P.R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Green J.L., Holder A.A.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Bayele H.K.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ489165; CAD32955.1; -.
FT NON_TER 1
FT NON_TER 504
SQ SEQUENCE 504 AA; 58810 MW; 514CC9110674F83A CRC64;

Query Match 8.3%; Score 125; DB 2; Length 504;
Best Local Similarity 21.5%; Pred. No. 8.8;
Matches 72; Conservative 60; Mismatches 109; Indels 94; Gaps 16;

Qy 3 EIVAD-----KTVE-VVKNAIETADGALDIY---NKYLDQVPMQTFDETIKELSRF 50
Db 1 EKVADINTYNEDPKEIEKKIENVVKKIDKKNIYKEINKLNISEIEKDKTSLEELKNI 60
Qy 51 KQESQAASVLVDGIKTLMDSDQKFEATQTVYEWCGVATQLLAAYILLFDEYNEKKAS 110
Db 61 NLSYGRS---LGNIFLEQIDEKKKAERT-----IKAMEAYIEDLDNKKKSD 106
Qy 111 AQKDILIKVLDDGITKLNEAQKSLVSSQSFNNASGK-----LLALDSOLTNDFS 160
Db 107 IEKDMKIKM-----DINEEMKALNISDDDRNVHTKSKDKKGGISDIHDKSKIIQNFS 160
Qy 161 EKS-----SYFQSQVDKIRK-----EAYAGAAAGVAVGPFGLIISYIAAGVVEGKLIP 209
Db 161 KESDINNKNELQENVSESRKNSDINHYSKVENI-----YNI-----1397
Qy 210 ELK-NKLKSVQNPFTTSLNTVQKANKDIDAAL---KLTTETI---AAIGIKTETETT-- 260
Db 200 -LKLNKIKKIIDKVKETDIEIKNNKNSINDELINSNGKIITKIKENSSLTECQSKIESTID 258
Qy 261 -----RFYVDYDDLMLSLKKEAAKMKINTCNEYOK 290
Db 259 DNYISKIKDIADLTKYILSEE-----NNINTYLYK 288

RESULT 18
Q7REV3 PRELIMINARY; PRT; 2749 AA.
AC Q7REV3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Hypothetical protein.
```

```
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE 235 kDa rhopty protein (Fragment).
GN Name-PY04930;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA Florens L., Yates P.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R., M.J.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "genome sequence and comparative analysis of the model rodent malaria
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001532; EAA16872.1; -.
DR InterPro; IPR0031900; KID repeat.
DR InterPro; IPR006499; ReticulocyteBP.
DR Pfam; PF02524; KID; 6.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 2749 AA; 322590 MW; PFE10079886F63CF CRC64;

Query Match 8.3%; Score 125; DB 2; Length 2749;
Best Local Similarity 21.5%; Pred. No. 64;
Matches 72; Conservative 60; Mismatches 109; Indels 94; Gaps 16;

Qy 3 EIVAD-----KTVE-VVKNAIETADGALDIY---NKYLDQVPMQTFDETIKELSRF 50
Db 1139 EKVADINTYNEDPKEIEKKIENVVKKIDKKNIYKEINKLNISEIEKDKTSLEELKNI 1198
Qy 51 KQESQAASVLVDGIKTLMDSDQKFEATQTVYEWCGVATQLLAAYILLFDEYNEKKAS 110
Db 1199 NLSYGRS---LGNIFLEQIDEKKKAERT-----IKAMEAYIEDLDNKKKSD 1244
Qy 111 AQKDILIKVLDDGITKLNEAQKSLVSSQSFNNASGK-----LLALDSOLTNDFS 160
Db 1245 IEKDMKIKM-----DINEEMKALNISDDDRNVHTKSKDKKGGISDIHDKSKIIQNFS 1298
Qy 161 EKS-----SYFQSQVDKIRK-----EAYAGAAAGVAVGPFGLIISYIAAGVVEGKLIP 209
Db 1299 KESDINNKNELQENVSESRKNSDINHYSKVENI-----YNI-----1337
Qy 210 ELK-NKLKSVQNPFTTSLNTVQKANKDIDAAL---KLTTETI---AAIGIKTETETT-- 260
Db 1338 -LKLNKIKKIIDKVKETDIEIKNNKNSINDELINSNGKIITKIKENSSLTECQSKIESTID 1396
Qy 261 -----RFYVDYDDLMLSLKKEAAKMKINTCNEYOK 290
Db 1397 DNYISKIKDIADLTKYILSEE-----NNINTYLYK 1426

RESULT 19
Q73KC5 PRELIMINARY; PRT; 1081 AA.
AC Q73KC5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
```







[illegible]

```
Qy 247 IAAIGBIKTETETTRFYVDYDDMLSL 273
Db 683 YIISGEV---VENTRYLIDVKNNQFTL 706

RESULT 24
CAE77235
ID CAE77235 PRELIMINARY; PRT; 713 AA.
AC CAE77235;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 13-APR-2004 (TrEMBLrel. 27, Last annotation update)
DE ATP dependant protease ClpB.
GN CLPB OR MSC_0613.
OS Mycoplasma mycoides (subsp. mycoides SC).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=44101;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PGI;
RC PubMed=14762060;
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundberg J.,
RA Johansson K.-E., Petersson B., Uhlen M.;
RA "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type
RT strain PGI, the causative agent of contagious bovine pleuropneumonia
RT (CBPP).";
RL Genome Res. 14:221-227(2004).
DR EMBL; BX842644; CAE77235.1; -.
SQ SEQUENCE 713 AA; 80782 MW; 1B5D204A9E29BE50 CRC64;

Query Match 8.2%; Score 123.5; DB 2; Length 713;
Best Local Similarity 22.3%; Pred. No. 16;
Matches 73; Conservative 52; Mismatches 131; Indels 71; Gaps 15;

Qy 2 TEIVADKTVVYVKNAIETADGALDLYNKYLDOVTPMQT-----FDETIKLSRP-- 50
Db 396 TGILVDRLLSSEKELNLELLKKYKVGQQAIVTSAIMRSRGKPNKDPKIGSFLF 455

Qy 51 -----KQEYSQA-ASVLVGDITKLLMDSQDKYFEATQTV-----YEWCGVAT 91
Db 456 LGPTGVGKTEVARSADILFNSPKMIRLDMSYNEKHSVAKLIGAPPYGVYEGGRLT 515

Qy 92 QLAA---YILFDEYNEKKAQAQDILIKVLDDG-----ITKLEAKSLLSVQSQFNN 143
Db 516 EAVRNPYSIVLFDRI-EKAHTDVFNILLQILDGRLTSLGKTIDFKNTIIVMTS--NI 572

Qy 144 ASGKLLALDSQLTNDFSEKSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYIAAGVV 203
Db 573 ASOYLTSDELQVVD-----DQKIQELNKKVFRPEFLNIDNIVY--FNALSQVTI--GEI 624

Qy 204 EGKLIPELKNKLSQNPFTLNTVKQANKDID-----AAKLK-----LTTE 246
Db 625 VDKVLEELSTRLDQDNQNFSEAR--NKLINEGYDLFCARPIKRYIEKNIETLIAH 682

Qy 247 IAAIGBIKTETETTRFYVDYDDMLSL 273
Db 683 YIISGEV---VENTRYLIDVKNNQFTL 706

RESULT 25
Q74DE2
ID Q74DE2 PRELIMINARY; PRT; 541 AA.
AC Q74DE2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN Name=hylB; ORFNames=GSU1374;
OS Geobacter sulfurreducens.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC Geobacteraceae; Geobacter.
OX NCBI_TaxID=35554;
```

```
RP SEQUENCE FROM N.A.
RX STRAIN=PCA / ATCC 51573;
RX PubMed=14671304; DOI=10.1126/science.1088727;
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,
RA Heidelberg J.F., Wu D., Wu M., Ward N.L., Beanan M.J., Dodson R.J.,
RA Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.I., Durkin A.S.,
RA Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,
RA Daviden J.M., Zafar N., White O., Tran B., Romero C., Forberger H.A.,
RA Weidman J.F., Khouri H.M., Feldblyum T.V., Utterback T.R.,
RA Van Aken S.E., Lovley D.R., Fraser C.M.;
RA "Genome of Geobacter sulfurreducens: metal reduction in subsurface
RT environments.";
RT Science 302:1967-1969(2003).
RL EMBL; AE017180; AAR34750.1; -.
DR TIGR; GSU1374; -.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR InterPro; IPR005829; Sug_transporter.
DR InterPro; IPR010989; t-snare.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS00885; HAMP; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN 1.
SQ SEQUENCE 541 AA; 57093 MW; 3C9BEDAEAF7C3B12 CRC64;

Query Match 8.2%; Score 123; DB 2; Length 541;
Best Local Similarity 22.7%; Pred. No. 13;
Matches 67; Conservative 44; Mismatches 112; Indels 72; Gaps 11;

Qy 15 NAETADGALDLYNKYLDOVTPMQTDETIKLSREFQESYQASVLYGDIKTLMDSQD 74
Db 32 NTARTANGLD--TVTRDVLV-----LKOLKIADMY----AVNVDVSHKVRNGNI 78

Qy 75 KYFEATQTVYEWCGVATQLLAAVILLFDEYNEKKAQAQDILIKVLDDGITKLEAKSL 134
Db 79 TWTGRKSVBEAKTKIAEKLAQVLAATNLAEEKKHLEAKPLIKVAD-----ATLER 130

Qy 135 LVSSQSFNNASGKLLALDSQLTNDFSEKSYFQSQVD--KIRKEAY----- 178
Db 131 LASILSAEDAEALTAFTVSELYPAIDPVSAKFSSIVDDQLKIAKQYDHSGLYRASRTI 190

Qy 179 --AGAAAGV-VAGPFGLLIISYIAAGVVE-----GKLIPE 210
Db 191 SLVAIVGVLIAGTAGLLITRSITGLAEGVEVANRLAAGDLTVEVRAGGRDETGMALAA 250

Qy 211 LKMKLSVQNFPT---TLSTNVQANKDIDAAKLTT---ETAA-IGETKTETE 258
Db 251 MGNMVTSLRLHIAEATISHSIGIASNQLHATSEQIATGSEVASQGVAVATASE 305

RESULT 26
AAR34750
ID AAR34750 PRELIMINARY; PRT; 541 AA.
AC AAR34750;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN HYL B OR GSU1374.
OS Geobacter sulfurreducens.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC Geobacteraceae; Geobacter.
OX NCBI_TaxID=35554;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PCA / ATCC 51573;
RX PubMed=14671304;
```

|    |  |  |
|----|--|--|
| RA | Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,        |  |
| RA | Heidelberg J.F., Wu D., Wu M., Ward N.C., Beanan M.J., Dodson R.J.,    |  |
| RA | Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,      |  |
| RA | Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,       |  |
| RA | Davidson T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A., |  |
| RA | Weidman J., Khouri H.M., Feldblyum T.V., Utterback T.R.,               |  |
| RA | Van Aken S.E., Lovley D.R., Fraser C.M.;                               |  |
| RA | "Genome of Geobacter sulfurreducens: metal reduction in subsurface     |  |
| RT | environments.";  |  |
| RT | Environ. Microbiol. Rev. 6:23-32 (2003).                               |  |
| RL | Science 302:1967-1969(2003).   |  |
| DR | EMBL; AE017211; AAR34750.1; --   |  |
| DR | TIGR; GSU1374; --  |  |
| SQ | SEQUENCE 541 AA; 57093 MW; 3C9BEDAEAF7C3812 CRC64;                     |  |

  

|  |                                    |
|--|------------------------------------|
| Query Match  | 8.2%; Score 123; DB 2; Length 541; |
| Best Local Similarity  | 22.7%; Pred. No.13;                |
| Matches 67; Conservative 44; Mismatches 112; Indels 72; Gaps 11; |                                    |

  

|    |     |   |     |
|----|-----|---|-----|
| QY | 15  | NAIETADGALDLYNKYLDVLPWQTFDETIKLSRFKQESQAASVLVGDITKLIMDSQD   | 74  |
| DB | 32  | NTARTANGLD--TVYDRVLV-----LKDILADMY-----AVNIVDVSHKVRNGNI     | 78  |
| QY | 75  | KYFEATQTVYVWCGVATQLLAAYILLFDEYNEKKASAKOILIKVLDDGITKLEAOKSL  | 134 |
| DB | 79  | TWTEGRKSVEAEAKTIAEKQLAYLATNLAEKKHLEAKPLIKVAD-----ATLER      | 130 |
| QY | 135 | LVSSQSFNNAAGKLLALDSQTNDFSEKSSVFSQOVD---KIRKEAY-----         | 178 |
| DB | 131 | LASTLSAEDAELTAFTVSELYPAIDPVSNAKFSLLVDDQLKIAQKQYDHSGLYRASRTI | 190 |
| QY | 179 | --AGAAAGV-VAGPGLIISVIAAGV-----GKLIPE                        | 210 |
| DB | 191 | SLVAIVGVLIAGTAGLITTSITGTLAEQVEVANRLAAGDLTVEVRAGGRDETQGLMAA  | 250 |
| QY | 211 | LKMKLKSQVQNPF--TSLNTVTKQAKDIDAACLKLT--EIAA-IGEIKTETE        | 258 |
| DB | 251 | MGNMVTSLRHLLAEAISISHGSIASNLQHTATSEQIATGTSSEVASQGVAVATASE    | 305 |

  

|           |  |
|-----------|--|
| RESULT 27 |  |
| Q9LVQ4    | PRELIMINARY; PRT; 649 AA.  |
| AC        | Q9LVQ4   |
| ID        | Q9LVQ4;  |
| DT        | 01-OCT-2000 (TREMBLrel. 15, Created)                                   |
| DT        | 01-OCT-2000 (TREMBLrel. 15, Last sequence update)                      |
| DT        | 01-OCT-2003 (TREMBLrel. 25, Last annotation update)                    |
| DE        | Myosin heavy chain-like.   |
| OS        | Arabidopsis thaliana (Mouse-ear cress).                                |
| OC        | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;     |
| OC        | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;   |
| OC        | eurosid II; Brassicales; Brassicaceae; Arabidopsiis.                   |
| OX        | NCBI_TaxID=3702;   |
| RN        | [1]  |
| RP        | SEQUENCE FROM N.A.   |
| RX        | MEDLINE=20181125; PubMed=10718197;                                     |
| RA        | Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,      |
| RA        | Tabata S.;   |
| RT        | "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence |
| RT        | features of the regions of 3,076,755 bp covered by sixty P1 and TAC    |
| RT        | clones.";  |
| RL        | DNA Res. 7:31-63(2000).  |
| DR        | EMBL; AB018120; BAA97285.1; --   |
| DR        | InterPro; IPR008545; DUF827.   |
| DR        | Pfam; PF05701; DUF827; 1.  |
| SQ        | SEQUENCE 649 AA; 72430 MW; B7D51EFB295D55BB CRC64;                     |

|                  |  |  |
|------------------|--|--|
| Qy               | 63   | GDIKTLMSDQKYFEATQTVYEWGVATQTLAAVILLFDEYNEK-----KA 107              |
| Db               | 180  | LETTVALSKVEEAKKYKSHSE-----KIELLRKEIAAVNVSVEQTKLACSOARKEQSEI 235    |
| Qy               | 110  | SAQKDILIKVLDDGITKLNEAQKSLLVSSQSFNNASGKLLALDLSOLTNDFFSEKSSYFQSQ 169 |
| Db               | 236  | FAEKEIQKSYKAG--MESSAKKSLAKNEFDPEFAK--KLEVQLTETYN----- 282          |
| Qy               | 170  | VDKIRKRAYAGAAAGVAGPPEGLIISYSIAAGVVEGKLIPELKNKLSVQNFFTLSTNV 229     |
| Db               | 283  | IDELQKQMET-AKASDIDSVNGVSLNELNEAKGLFE-KLVEBEKSLQELVESLKAEKLN-V 339  |
| Qy               | 230  | KQANKDIDAALKLTLTEIA-----AIGEIKTE-----TETTFYVDYDDILMELSLK----- 275  |
| Db               | 340  | KMEHDEVEAKEAETIESVAGDILHLKLSRKSSELCQCVTBESKAAALEDDMLTLNQTISSET 399 |
| Qy               | 276  | EAAKK-----MINTCNEYOK 290   |
| Db               | 400  | EAAREAGRWKAKELMK 418   |
| RESULT 28        |  |  |
| Q8CP10           |  |  |
| ID               | Q8CP10   | PRELIMINARY; PRT: 1189 AA.   |
| AC               | Q8CP10;  |  |
| DT               | 01-MAR-2003  | (TrEMBLrel. 23, Created)   |
| DT               | 01-MAR-2003  | (TrEMBLrel. 23, Last sequence update)                              |
| DT               | 01-MAR-2004  | (TrEMBLrel. 26, Last annotation update)                            |
| DE               |  | Chromosome segregation SMC protein.                                |
| GN               |  | OrderedLocusNames=SE0909;  |
| OS               |  | Staphylococcus epidermidis.  |
| OC               |  | Bacteria; Firmicutes; Bacillales; Staphylococcus.                  |
| NCBI_TaxID=1282; |  |  |
| [1]              |  |  |
| RP               |  | SEQUENCE FROM N.A.   |
| RC               |  | STRAIN=ATCC 12228;   |
| RX               |  | PubMed=12950922;   |
| RA               | Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,            |  |
| RA               | Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,         |  |
| RA               | Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;                    |  |
| RT               |  | "Genome-based analysis of virulence genes in a non-biofilm-forming |
| RT               |  | Staphylococcus epidermidis strain (ATCC 12228).";                  |
| RL               | Mol. Microbiol. 49:1577-1593(2003).                                      |  |
| RE               | EMBL; AB016747; AAC04506.1; -  |  |
| DR               | HSP; P58301; 1F2T.   |  |
| DR               | GO; GO:0016020; C:membrane; IEA.   |  |
| DR               | GO; GO:0005634; C:nucleus; IEA.  |  |
| DR               | GO; GO:0005524; F:ATP binding; IEA.                                      |  |
| DR               | GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA. |  |
| DR               | GO; GO:0005525; F:GTP binding; IEA.                                      |  |
| DR               | GO; GO:0007059; P:chromosome segregation; IEA.                           |  |
| DR               | GO; GO:0006810; P:transport; IEA.  |  |
| DR               | InterPro; IPR003439; ABC transporter.                                    |  |
| DR               | InterPro; IPR005289; GTP-binding.  |  |
| DR               | InterPro; IPR003405; SMC_C.  |  |
| DR               | InterPro; IPR010935; SMC_hinge.  |  |
| DR               | Pfam; PF02483; SMC_C; 1.   |  |
| DR               | Pfam; PF06470; SMC_hinge; 1.   |  |
| DR               | Pfam; PF02463; SMC_N; 1.   |  |
| DR               | TIGRFAMs; TIGR00650; MG442; 1.   |  |
| DR               | Complete proteome.   |  |
| Q8CP10           |  | SEQUENCE 1189 AA; 137542 MW; 0F6F95636BE31BC8 CRC64;               |
| Q8CP10           |  |  |

```

Query Match      8.2%; Score 123; DB 2; Length 1189;
. Best Local Similarity 19.6%; Pred. No. 32;
Matches 70; Conservative 52; Mismatches 112; Indels 124; Gaps 13;

Qy 5 VADKTVEVK-----NAISTAGALDLYNKYLDQVIPWQTFDETIELSLRFKOEYSQAA 58
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 161 IIESAGLVLYKKRKAESIQKLDHTDNLRVDEILDYDLEGRVPLKEEAIAIKYKQLS 220

```

```
Qy 59 S-----VLVDGIKTLMSOD-----KYFEATQTVYEWCGVATQALLAAYILLFDEYN 105
Db 221 KEMEQSVDIVTVSDIDHYEDNQDLRLNHLKSQAQKE--GQAQINQ-----LLQKYK 274
Qy 106 EKXASAKDILIKVLDDGITKLANEAKSLVSSQSFNNASGKLLALDSQLTND-----F 159
Db 275 GKQONDYD-----IEKLN--VELVKATENYEQLSGKLVLEERKKNQSETNARY 322
Qy 160 SEKSYFSQSQVDKIRKAYAGAAAGVVGPFGLIISYSIAAGVVEGKLPKLN-----213
Db 323 EEELDNLESQIDSINKNEK-----AQNEKLLAELAKNKKQQLN 358
Qy 214 -----KLKSVONFPTTLNNTVKQAKNDI-----DAAKLKLT 244
Db 359 KEVQESLSLYISDEQHEKLEEFKNSYYTLMSEQSDVNNDIRPLEHTINENAKSRLD 418
Qy 245 TEIAAIGEIKETETTFYVDYDDLMLSLKKEAAKKNMTCNEYQKRHGKTLFVPE 302
Db 419 SRLVE-----AFNLKDIQONITQKQYQS--SKKSMKVRQ 454

RESULT 29
Q9BJY0 PRELIMINARY; PRT; 2752 AA.
AC Q9BJY0;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE 235 kDa rhothry protein (Fragment).
GN Name=Illa.1;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21273138; PubMed=11378199;
RA Khan S.M., Jarra W., Bayele H., Preiser P.R.;
RT "Distribution and characterisation of the 235 kDa rhothry multigene family within the genomes of virulent and avirulent lines of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 114:197-208(2001).
DR EMBL; AF233442; AAAK5625.1; -.
DR InterPro; IPR006499; ReticulocyteBP.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
FT NON_TER 1
FT NON_TER 2752
FT NON_TER 2752
SQ SEQUENCE 2752 AA; 322608 MW; E834E3753168AF76 CRC64;

Query Match 8.2%; Score 123; DB 2; Length 2752;
Best Local Similarity 21.3%; Pred. No. 86;
Matches 72; Conservative 58; Mismatches 108; Indels 100; Gaps 16;

Qy 3 EIVAD-----KTYE-VVKNAIETADGALDY--NKYLDQVIPWOTFDETIKELSRF 50
Db 1140 EKVADINTYXEPKEIEKKIENVVKKKKKIYKINKLNEISEIEKDKTSLBELXNI 1199
Qy 51 KQYSOAAVLVGDIKTLMSODKQYFEATQTVYEWCGVATQALLAAYILLFDEYNEKAS 110
Db 1200 NLSYGRS-----LGNIFLEQIDEKKAERT-----IKAMEAYIEDLDNIKKSD 1245
Qy 111 AQKDILIKVLDDGITKLANEAKSLVSSQSFNNASCK-----LLALDSQLTNDPS 160
Db 1246 IEKDMKIKM-----DINEEMKALNISDDDRNYHTKSKDKHGKSIDHDKSKKIQNPS 1299
Qy 161 EKS-----SYFOSQVDKIRK-----EAYAGAAGVVGPFGLIISYSIAAGVVEGKLP 209
Db 1300 KESDINNKNELQENYSESKHNSDINHLYSKVENI-----YNI-----1338
Qy 210 ELK-NKLKSVONFPTTLNNTVKQAKNDI-----RAKULTTEIAAIGEIKETETTF 259
Db 1339 -LKLNKIKIIDVKVEYTDIEIKNNKSINDELXNSGKILTKIK---ENSLTECOSKIES 1394
Qy 260 T-----RFVVDYDDLMLSLKKEAAKKNMTCNEYQKRHG 290
```

```
Db 1395 TIDDNYSIKCIKDIADLKTYILSEE-----NNINTYLK 1427

RESULT 30
Q71X69 PRELIMINARY; PRT; 927 AA.
AC Q71X69;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
DE Membrane protein, putative.
GN Ordered locus names=LMOF2365_2330;
OS Listeria monocytogenes (serotype 4b / strain F2365).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=265669;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15115801; DOI=10.1093/nar/gkh562;
RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,
RA Kolonay J.F., Rasko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,
RA Peterson J.D., White O., Nelson W.C., Niernan W.C., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,
RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,
RA Forberger H.A., Tran B., Katharion S., Wondolung L.D., Uhlich G.A.,
RA Bayles D.O., Luchansky J.B., Fraser C.M.;
RT "Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen Listeria monocytogenes reveal new insights into the core genome components of this species.";
RL Nucleic Acids Res. 32:2386-2395(2004).
DR EMBL; AE017330; AAT05096.1; -.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR001809; Outrsurface.
DR Pfam; PF00015; MCPsignal; 1.
DR ProDom; PD001127; Outrsurface; 1.
KW Complete proteome.
SQ SEQUENCE 927 AA; 97232 MW; 9AE73C67B14685CA CRC64;

Query Match 8.1%; Score 122.5; DB 2; Length 927;
Best Local Similarity 21.7%; Pred. No. 26;
Matches 75; Conservative 35; Mismatches 137; Indels 99; Gaps 12;

Qy 3 EIVADKTVVEVKNAIETADGALDLYNKYLDQVIPWOTFDETI-----KELSRFKQEVSOAA 58
Db 166 EAIFDKIKESGEGFAQAADGS-----GKIKDGLVKSQEGNKTISTNLKTLADSLTFKQGA 221
Qy 59 SVL-----VGDIK-----TLMSODKQYF 77
Db 222 NTLLEVGLKTYTDGVNTAAAGGDKLNDGVSTLAAGVGLKDGVAALDGGATKLASGVSTYT 281
Qy 78 EATQTVYEWCGVATQALLAAYILLFDEYNEKKAQAQDKIILIKVLDDGITKLANEAKSLVLS 137
Db 282 SGVDTL---AGGINQATYGTSTALSDGLNKNKS-----VPTLASGITQLNNGQKSLATG 332
Qy 138 SQSFNNASGKLLA-----LDSQLTN-----DFSEKSYFSQSQVDKIRKAYAGAAGVVA 187
Db 333 LDSLVDSGNSKLKSLAGLKDGLNLTDKQKIAQLKQGMNDLQOGIDOLNKSXVNGBDAA---- 388
Qy 188 GPFGLIISYSIAAGVVEGKLPKLNKLVKSVQNFPTTLNNTVKQAKNDIADAKLKLTETI 247
Db 389 -----LAKOLAA-----LQKSLDGLQNLGTLFIKS---NANFDAEAKSKINATA 429
Qy 248 AAIGEIKETETTFYVDYDDLMLSLKKEAAKKNMTCNEYQKRHG 293
Db 430 GVSADKQKI-----IDAIQADLDKETQKTSATQVATVEQLQSG 467

RESULT 31
AAT05096 PRELIMINARY; PRT; 927 AA.
AC AAT05096;
DT 10-MAY-2004 (TREMELrel. 27, Created)
DT 10-MAY-2004 (TREMELrel. 27, Last sequence update)
```

DT 10-MAY-2004 (TReMBLrel. 27, Last annotation update)  
DE Membrane protein, putative.  
GN LMOF2365\_2330.  
OS Listeria monocytogenes str. 4b F2365.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria;  
OC Listeria monocytogenes.  
OX NCBI\_TaxID=265669;  
EN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=4b F2365;  
RX PubMed=15115801;  
RA Nelson K.E., Fouts D.E., Mongodin E.P., Ravel J., DeBoy R.T.,  
RA Kolonay J.F., Raeko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,  
RA Peterson J., White O., Nelson W.C., Niernan W., Beanan M.J.,  
RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,  
RA Hatt D.H., Selengut J., Van Aken S., Khouri H., Fedorova N.,  
RA Forberger H., Tran B., Katharopoulos S., Wondolting L.D., Unlich G.A.,  
RA Bayles D.O., Luchansky J.B., Fraser C.M.;  
RA "Whole genome comparisons of serotype 4b and 1/2a strains of the food-  
RT borne pathogen *Listeria monocytogenes* reveal new insights into the  
RT core genome components of this species."  
RL Nucleic Acids Res. 32:2386-2395(2004).  
DR EMBL; AE017330; AAT05096.1; -;  
SQ SEQUENCE 927 AA; 97232 MW; 9AB73C67B14685CA CRC64;

Query Match 8.1%; Score 122.5; DB 2; Length 927;  
Best Local Similarity 21.7%; Pred. No. 26;  
Matches 75; Conservative 35; Mismatches 137; Indels 99; Gaps 12;  
QY 3 EIVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETI-----KELSRFKQEQYQAA 58  
DB 166 EAIKFKIKESGEGFAAGDS-----GKIKGLGVKSQGNKTIISTNLKTLADSSLTFFKQGA 221  
QY 59 SVL-----VGDIK-----TLLMDSQDKYF 77  
DB 222 NTLVGLKTYTDGVTNAAAGDKLNDGVSTLAAGVPLKDGVAALDGGATKLAGSVTYT 281  
QY 78 EATQTVVWCQVATQLAAAYILLDFEYNEKASQAKDILIKVLDDGTTKLENAQKSLVLS 137  
DB 282 SGVDTL---AGGINQAVTGSTALSDGLNKNMGS-----VPTLASGITQLNNGQKSLATG 332  
QY 138 SQSFNNASGKLLA---LDSOLTW-----DFSEKSSVFQSQVDKIRKEAYAGAAAGVVA 187  
DB 333 LDSLVDSGNKLSAGLKGELDGLNLTQKQKIAQLKGMDLQGIQDILNKNVNGEDAA----- 388  
QY 188 GPFGLIISYSIAAGVVEGLKIPELNKLKLSVQNFETTLSTNTVQANKDIDAAKLKLATTEI 247  
DB 389 -----LAKQLAA-----LQKSLGDLQNGLTFIKS---NANFDAEAIKSKINATA 429  
QY 248 AAIQEIKTETETTFYVDYDMLSLKLEAKAKRMINTCNEYQKRHG 293  
DB 430 GVSADKQKI-----IDAIQADLDKETQKTSATQVATVEQLQSG 467

RESULT 32  
Q7SH24  
ID Q7SH24 PRELIMINARY; PRT; 4007 AA.  
AC Q7SH24;  
DT 01-MAR-2004 (TReMBLrel. 26, Created)  
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE Hypothetical protein.  
GN Name=NCU00658.1;  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
EN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=OR74A.  
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,  
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Reiman B.,  
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,

RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,  
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,  
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,  
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,  
RA Kamal M., Kamvysellis M., Mauceli E., Bielke C., Rudd S., Frishman D.,  
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,  
RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmari S.A.,  
RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,  
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,  
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,  
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.;  
RT "The Genome Sequence of the Filamentous Fungus *Neurospora crassa*."  
RL Nature 0:0-0(2003).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC EMBL; AABX01000001; EAA36562.1; -;  
DR InterPro; IPR009638; Fez1.  
DR InterPro; IPR003900; KID repeat.  
DR InterPro; IPR001638; SBP\_bac\_3.  
DR Pfam; PF06818; Fez1; 1.  
DR Pfam; PF02524; KID; 5.  
DR PROSITE; PS01039; SBP\_BACTERIAL\_3; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 4007 AA; 453246 MW; 67CDF6E6F1463612 CRC64;

Query Match 8.1%; Score 122.5; DB 2; Length 4007;  
Best Local Similarity 21.4%; Pred. No. 1.4e+02;  
Matches 73; Conservative 57; Mismatches 132; Indels 79; Gaps 13;  
QY 11 EVKNAIETADGAL-----DLYNKYLDQVWPQTFDETIK-----ELSRFKQEQYQ 56  
DB 1159 EILKQHQSVGESESIATIKYKDKDLSRNTSQAIDKLQHENELANFKAYEQ 1218  
QY 57 AASVLGVDIITLMDSDQKFEATQTVVYCGVATQLAAVILLDFEYNEKKA----- 109  
DB 1219 EKQLAVQHKTEMSLTDRVHEKEK-----LATQYQERVQALSAELADKKTALAEYKE 1271  
QY 110 --SAQKIDILIKVLDDGTTKLENAQKSL-----LVSSQSFNNAS-----GKLALD--- 152  
DB 1272 QLSASRAQLDKLADHGVKVDLQAKLSEVAKVTADYEGNLSLRTKHQGEVNVLVHH 1331  
QY 153 ----SQTNDPSEKSSVFQSQVDKIRKEAYAGAAAGVAGPFGILLIISYIAAGVVEGLK- 207  
DB 1332 QDEIKKLTAGNEKIRNLEHRIINDLKAELKQDRA-----BFDKKKALLEGEVA 1379  
QY 208 ----IPELNKLKLSVQNFETTLSTN---TVQANKDIDAAKLKLATTEIAAIGIKTETE 258  
DB 1380 TLQGVDDKSKSLSSKEAEFELKLENAQIAELRKDVADKSNLSQDKLELSDLKGQOK 1439  
QY 259 T--TRFYVDYDMLSLK-----EAAKMIN-TCNEYQKR 291  
DB 1440 TRIEDFNVQINERMAQLKKAQNELKASQASLNTTTTTEYDAK 1480

RESULT 33  
Q8I754  
ID Q8I754 PRELIMINARY; PRT; 852 AA.  
AC Q8I754;  
DT 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Rhostry protein (Fragment).  
GN Name=rhop235-8;  
OS Plasmodium yoelii yoelii.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=73239;  
EN [1]  
RN SEQUENCE FROM N.A.  
RA Khan S.M., Bayele H., Preiser P.R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.

```

RA Green J.L., Holder A.A.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Baye H.K.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ489163; CAB32953.1; -.
FT NON TER 1
FT SEQUENCE 852 AA; 99060 MW; 026995B7A6B658B1 CRC64;
SQ SEQUENCE 852 AA; 99060 MW; 026995B7A6B658B1 CRC64;

Query Match 8.1%; Score 122; DB 2; Length 852;
Best Local Similarity 21.1%; Pred. No. 25;
Matches 71; Conservative 61; Mismatches 108; Indels 96; Gaps 16;

Qy 3 EIVAD-----KIVE-VVNALETADGALDLY---NKVLDQVIPWQTFDETIKLSRF 50
Db 1 EKVDADINYNEDPEIEKKIENVVKKIDKKNIYKEINKLLNEISEIEKDKTSLLELNKI 60
Qy 51 KQEYSQASVLVGDITKLLMDSQDKYFEATQTVYEMCGVATQLLAAAYLLFDEYNEKAS 110
Db 61 NLSYGRS---LGNIFLEQIDEKKKAERT-----IKAMEAYIEDLDNIKKKSDE 106
Qy 111 AQKDILIKVLDDGITKLNBAQKSLVSSQSFNNASGKLLALDSQL--TNDFSEKSYF--- 166
Db 107 IEKDMKIKM-----DINEEMKALNINSDDRNYHTK--SKDHKKGISDIHDKSSKIION 158
Qy 167 --OSQVDKIRKEAYAGAAAGVAGPGLIISYSIAAGVVEG-KLIPELKNKLSKVQNF 222
Db 159 FSKESDINNITE-----LQENVSESKHNSDINHLSKVENIY 197
Qy 223 TTL-----SNTVKQANKDIDAACL---KLTFEI---AAIEGKTTETTT- 260
Db 198 NILKLNKIKKIIDVKVEYTDIKKNNKSINDELINSKGIITIKENSLSLTCQSKIESTI 257
Qy 261 -----RFYVDYDILMLSLKKEAKKMNINTCNEYQK 290
Db 258 DDNYISKICIKDADLKTYLSE-----NNINTYLK 288

RESULT 34
Q9SZK7 PRELIMINARY; PRT; 1496 AA.
AC Q9SZK7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein F20D10.190 (Hypothetical protein AT4g38070).
GN Name=F20D10.190; Synonymus=AT4g38070;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA Bevan M., Wedler H., Kutzner M., Wambutt R., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Schueller C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Wedler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035538; CAB37547.1; -.

```

```

DR EMBL; AL161592; CAB80472.1; -.
DR FIR; T05634; T05634. HLH_basic.
DR InterPro; IPR001092; HLH_1.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS50888; HLH; 1.
KW Hypothetical protein.
SQ SEQUENCE 1496 AA; 173346 MW; 03BBB75DCD2B03FE CRC64;

Query Match 8.1%; Score 122; DB 2; Length 1496;
Best Local Similarity 18.9%; Pred. No. 49;
Matches 64; Conservative 72; Mismatches 124; Indels 78; Gaps 14;

Qy 4 IVADKTVVVKNAIETADGALDLYNK--YLDQVIPWQTF-----FDETIKLSRF---KQ 52
Db 153 IEAEKTVKGMKGRDRDDVVVKMEESQVVEEKWKKEQFKHLEAYEKLKNLFKDSKK 212
Qy 53 EYQASASVLVGDITKLLMDSQDKYFEATQTVYEMC-GVATQLLAAAYLLFDEYNEKKA 109
Db 213 EWBEESKLLDEIYSIQTKLDSVTRISEDQKKLQMCNGALTO-----BETRRKHL 263
Qy 110 SAQKDILIKVLDDGITKLNBAQKSLVSSQSFNNASGKLLALDSQLTNDFSEKSYFQ-- 167
Db 264 EIQVSEFKAYVEDAFACQDARTQL-----DDLAKGRDWEVAELRQTLMSKDAYFKEM 316
Qy 168 ----SQVDKIRKEAYAGAAAGVAGPGLIISYSIAAGVVEG---KLIPELKNKLSKVON 220
Db 317 KYENGKLEQENRELLGSLK-----ELQEATIQSGNSALSCLKNKNFRLEN 362
Qy 221 FFTTL-----SNTVKQANKDIDAACLKLITTEIAAIGIKTETE-----TTRFYV 264
Db 363 IHKNCANLSKAEAWSSQVEKWEINQYKLOQSKAAALKEVELENCRSSTAKMRL 422
Qy 265 DYDDLMLSL-----KEAAKMNINTCNEY---QKRGH 293
Db 423 QYEEISIMFVLVSRVTVSEAQSRLANAKDKQIKDEKREG 460

RESULT 35
Q7RFQ7 PRELIMINARY; PRT; 1081 AA.
AC Q7RFQ7;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 235 kDa rhostry protein (fragment).
GN Name=PY04647;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
[1]
RP SEQUENCE FROM N.A.
RA STRAIN=17XNL;
RA PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Koolij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallon S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoaihi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., O.R.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002)
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001425; EAA16541.1; -.
FT NON TER 1
SQ SEQUENCE 1081 AA; 126062 MW; DDFFE8EFDD6D6A94 CRC64;

```

Query Match 8.1%; Score 121.5; DB 2; Length 1081;  
 Best Local Similarity 19.1%; Pred. No. 36;  
 Matches 74; Conservative 72; Mismatches 129; Indels 113; Gaps 17;

QY 4 IVADKTVVVKNAIETAD-----GALDLYNKYLDOVIPWOTFDTETIKELSRFKQY 54  
 DB 589 LVACSTLINENIEAQNDEQIKKHLGELDSKVDKINI-----TFDEISKVLNNVNNK- 643

QY 55 SQAASVLVGDIKTL-----LMSDQKYPEATQTVVWCGVATQLLAAAYILLFD- 102  
 DB 644 KEALLELINEIKTYVYHMINDLSSILNSVQTIDKH-----LQSSIDSYGLYNELKNYKEE 700

QY 103 -----EYNEKKA-----SAQDILIKVLDDGKITKLEAKQSLIYS-----SQSF 141  
 DB 701 ILKKNMDYEDYKSKILEYNNFISKENIILKIDIEE-ISKISNEVKTSLPKYDNEVKYIY 759

QY 142 NNASGKLLALDSOL-----TNDFSEKSSYFOSQVDKIRKEAYAGAAA 183  
 DB 760 PNSNGEYLVQVKLIEQIKETYSBYLNKYRNKNDIKTVSDTKQIQTNRKMDNYKVL 819

QY 184 GVVAGPGLIISYIAAGVVEGK--LPELKNKLKSVQNFPTLSNTVQOA-----NKDI 236  
 DB 820 NTVMQVQSI--YELIKNIINGKNVLDLLRQSIQAION-YDNINETVKNSSITELNNKI 876

QY 237 DAAKLLKTTTIAAIGETETETRFY-----VDYDDLMSL----- 273  
 DB 877 SMIDDKLNT--TTINDIETKLSTLKEYFTSTKQIOMENSEIEDNLLQNVNEIDOVQNST 934

QY 274 -----LKEAKKMINTCNEVQKH 292  
 DB 935 NALNEQYISLMDVEKLINDTNEIRKH 962

RESULT 36  
 Q7RAW8 PRELIMINARY; PRT; 1116 AA.

AC Q7RAW8  
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE P235 rhotry protein B5 (Fragment).  
 GN Name=PY06381;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=73239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17XNI;  
 RX PubMed=12368865;  
 RA Carlton J.M., Angiuoli S.V., Suh B.B., Koolij T.W., Pertea M.,  
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallow S.J., van Aken S.E., Riedmuller S.B., Feidblyum T.V.,  
 RA Chao J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,  
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;  
 RT "Genome sequence and comparative analysis of the model rodent malaria  
 RT parasite Plasmodium yoelii yoelii.";  
 RL Nature 419:512-519(2002).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC EMBL; AAL01002149; EAA18583.1; -.  
 FT NON TER 1116 1116  
 SQ SEQUENCE 1116 AA; 131905 MW; 3B264B3B5701B222 CRC64;

Query Match 8.0%; Score 121; DB 2; Length 1116;  
 Best Local Similarity 21.0%; Pred. No. 40;  
 Matches 72; Conservative 57; Mismatches 134; Indels 80; Gaps 17;

QY 7 DKTVEVVVKNAIETADGALDLYNKYLDOVIP-WOTFDTETIKELSRFKQYEQYSAASVLVGDI 65  
 DB 460 ENTLDYFNKK-RNRDRNSNDI-NKNILLIPLMDQFDTLKNKSMKLKNDAASEKYVIITQI 517

QY 66 KTLMDSD--QDK---YFEATQTVVWCGVATQLLAAAYILLFDYNEKKAQAQDI--LIK 118  
 DB 518 KQKLNOSTYDDKKTGTFSALKAESW---ETKELETIAKLNK-NEETVVKLENEIRDLFK 573

QY 119 VLDDGIT-----KLNEAQKSLIYSQSNNASGKLLALDSOLTN----- 157  
 DB 574 KYSDEVAEKYIIEELKCLKETIKDIYKKYK--RAIDLKKEIENKNKYIDELGKK 629

QY 158 -----DPEKSSYFOSQVDKIRKEAYAGAAVWGPGLIISYIAAGVWEGKLIPEL 211  
 DB 630 TPFQIBEYVKKDTIYSTIKTELSEYK-----NIELYNELSSVQENTIDPI 679

QY 212 KKK-----LKS-VQNFPTLSN---TVQANKDIDAALKLTETAAI-----GEIKTE 256  
 DB 680 KKKKLETLKSEIDNVYKIQNKIEVELHLKNIETSKNELSNLILIIKIYFGEIDND 739

QY 257 TETTRFYVDYDDLMSLLEAKKMINTCNEVQKHGKTKLFE 299  
 DB 740 LNKT-----LKEFKNKEQELSNKINDYTKENDQLSVYQ 772

RESULT 37  
 Q8N592 PRELIMINARY; PRT; 961 AA.

AC Q8N592  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Vesicle docking protein pl15.  
 GN Name=VDP;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalek U., Smallos D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Strausberg R.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC032654; AAH32654.1; -.  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008565; F:protein transporter activity; IEA.  
 DR GO; GO:0006886; P:intracellular protein transport; IEA.



```

DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR InterPro; IPR006955; Uoi1_p115_C.
DR InterPro; IPR006953; Uoi1_p115_head.
DR Pfam; PF04871; Uoi1_p115_C; 1.
DR Pfam; PF04869; Uoi1_p115_head; 1.
DR PROSITE; PS50176; ARM_REPEAT; 1.
SQ SEQUENCE 961 AA; 107765 MW; DCF4888F4B8C02A CRC64;

Query Match      8.0%; Score 120.5; DB 2; Length 961;
Best Local Similarity 23.7%; Pred. No. 36;
Matches 74; Conservative 51; Mismatches 136; Indels 51; Gaps 13;

QY 7 DKTVEVVKNAIETADGALDLY-NKYLQVPIPWOTFDTIKELGRFKQYVQAASVVLGDI 65
Db 634 DKKEEVKTKLEQHDNIVTHYKNMIREQDLQLELRQOVSTLKQNEQLQTAQTAVTQVSOI 693
QY 66 K-----TLIM-----DSQDKYFEATQVYVCGVATQLLAAVILLFDEYNEKKAS-- 110
Db 694 QQHKDQYNLLKIQLGKDNQHGYSYSEGAQ-----MNGIQPEEIGR---LREEIELKRNQE 746
QY 111 -----AQKDILIKVLDDGITKLENAQKSLIVS---SQSFNNASGKLLALDLSQLTNDPSE 161
Db 747 LLOSQLTEKDSMIENKSSQTSCTGNESSAIVSARDSEQVAELKQELATLKSQV-NSQSV 805
QY 162 KSSYFQSQVDKIRKEAYAGAAAGVAGPFGGLIISYSIAAGVVEGKL-----IPELKNKL 215
Db 806 EITKLQTEKQELLQKTEAFKSVVEQGETETIATKTDD--VEGRLSALLQETKELKNEI 863
QY 216 KSVQNFPTLNTVKQANKDIDAAKLKLTTETIAAIGIKETETTRFYVDYDDLMLSLLK 275
Db 864 KALSEERTAIKEQLDSSNSTIAI-----LQTE-----KDKLEITDSKKQDQLLV-LLA 913
QY 276 EAAKQKMTNCNE 287
Db 914 DQDQKILSLKNK 925

RESULT 38
VDP_HUMAN
ID VDP_HUMAN STANDARD; PRT; 962 AA.
AC O60763;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE General vesicular transport factor p115 (Transcytosis associated protein) (TAP) (Vesicle docking protein).
DE Name=VDP;
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PHOSPHORYLATION SITE SER-942.
RX MEDLINE=98148093; PubMed=9478999;
RA Sotda M., Misumi Y., Yano A., Takami N., Ikehara Y.;
RT "Phosphorylation of the vesicle docking protein p115 regulates its association with the Golgi membrane."
RL J. Biol. Chem. 273:5385-5388(1998).
CC -!- FUNCTION: General vesicular transport factor required for intercellular transport in the Golgi stack; it is required for transcytotic fusion and/or subsequent binding of the vesicles to the target membrane. May well act as a vesicular anchor by interacting with the target membrane and holding the vesicular and target membranes in proximity (By similarity).
CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein which recycles between the cytosol and the Golgi apparatus during interphase.
CC -!- DOMAIN: Composed of a globular head, an elongated tail (coiled-coil) and a highly acidic C-terminal domain.
CC -!- PTM: Phosphorylated in a cell cycle-specific manner; phosphorylated in interphase but not in mitotic cells.
CC Dephosphorylated protein associates with the Golgi membrane; phosphorylation promotes dissociation.

```

```

CC -!- SIMILARITY: Belongs to the VDP/Uoi1/YBL047C family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D86326; BAA25300.1; -.
DR MIM; G03344; -.
DR GO; GO:0000139; C:Golgi membrane; TAS.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR InterPro; IPR006955; Uoi1_p115_C.
DR InterPro; IPR006953; Uoi1_p115_head.
DR Pfam; PF04871; Uoi1_p115_C; 1.
DR Pfam; PF04869; Uoi1_p115_head; 1.
DR PROSITE; PS50176; ARM_REPEAT; UNKNOWN_1.
KW Coiled coil; Golgi stack; Membrane; Phosphorylation; Protein transport; Transport.
FT DOMAIN 1 637 Globular head.
FT DOMAIN 638 930 Coiled coil (Potential).
FT DOMAIN 935 962 Asp/Glu-rich (acidic).
FT MOD_RES 942 942 Phosphoserine.
FT MUTAGEN 942 942 S->A: Loss of phosphorylation.
SQ SEQUENCE 962 AA; 107906 MW; 2E748F2C1BC2B942 CRC64;

Query Match      8.0%; Score 120.5; DB 1; Length 962;
Best Local Similarity 23.7%; Pred. No. 36;
Matches 74; Conservative 51; Mismatches 136; Indels 51; Gaps 13;

QY 7 DKTVEVVKNAIETADGALDLY-NKYLQVPIPWOTFDTIKELGRFKQYVQAASVVLGDI 65
Db 635 DKKEEVKTKLEQHDNIVTHYKNMIREQDLQLELRQOVSTLKQNEQLQTAQTAVTQVSOI 694
QY 66 K-----TLIM-----DSQDKYFEATQVYVCGVATQLLAAVILLFDEYNEKKAS-- 110
Db 695 QQHKDQYNLLKIQLGKDNQHGYSYSEGAQ-----MNGIQPEEIGR---LREEIELKRNQE 747
QY 111 -----AQKDILIKVLDDGITKLENAQKSLIVS---SQSFNNASGKLLALDLSQLTNDPSE 161
Db 748 LLOSQLTEKDSMIENKSSQTSCTGNESSAIVSARDSEQVAELKQELATLKSQV-NSQSV 806
QY 162 KSSYFQSQVDKIRKEAYAGAAAGVAGPFGGLIISYSIAAGVVEGKL-----IPELKNKL 215
Db 807 EITKLQTEKQELLQKTEAFKSVVEQGETETIATKTDD--VEGRLSALLQETKELKNEI 864
QY 216 KSVQNFPTLNTVKQANKDIDAAKLKLTTETIAAIGIKETETTRFYVDYDDLMLSLLK 275
Db 865 KALSEERTAIKEQLDSSNSTIAI-----LQTE-----KDKLEITDSKKQDQLLV-LLA 914
QY 276 EAAKQKMTNCNE 287
Db 915 DQDQKILSLKNK 926

RESULT 39
Q7RD43
ID Q7RD43 PRELIMINARY; PRT; 1611 AA.
AC Q7RD43;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY05582;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;

```







```
Db 1363 AKTELSDLITSSNQAPADAIS-----DAQKTLDEI 1393
RESULT 43
CAF32691 PRELIMINARY; PRT; 1828 AA.
AC CAF32691
DT 31-MAR-2004 (TrEMBLrel. 27, Created)
DT 31-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 31-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Variable membrane protein precursor.
GN VMP.
OS Mycoplasma hominis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4195;
RA Boesen T.;
FT SIGNAL 1 28 Potential.
FT CHAIN 29 1828 variable membrane protein.
SQ SEQUENCE 1828 AA; 205343 MW; 6A0D98F6491A1570 CRC64;
Query Match 8.0%; Score 120; DB 2; Length 1828;
Best Local Similarity 19.8%; Pred. No. 82;
Matches 79; Conservative 53; Mismatches 111; Indels 156; Gaps 14;
QY 2 TEIVADKTEVVKNAITAGAL-----DLNKKYLD-----QVLPWQTFD 41
Db 1051 TKALEDKSELAQNIIDTKNNATKTFNDKKDELNKLDTSDAKSVNKKESDVLKONSID 1110
QY 42 ET--IKELSPKQEYSOASVIVGDIKTLMDSDQKYFEATQTYEHCQVATQLLAAYIL 99
Db 1111 SNTPIKEIKATEKITEAINSLTTSIK-----NKKDEE----- 1143
QY 100 LFDEYNKKSAQKDIILIKVLDDGITKLNEAQKSLV-----SSQSFNNAS 145
Db 1144 -FNKYNDIKTSLEN--LIKEEDAVQVGIADVQKTLSENNVDKATIEKIQHSTEALTHAK 1200
QY 146 GKLLAL-----DSQLTNDFSEKSSYFQ-----SQVDKIRKEAYAGAAAGVAGPPGLI 193
Db 1201 EELKKLIDTTKKQLTKEFETKSELEKLSLIPDANNVDKKDELISFG----- 1247
QY 194 ISYSIAGVVEGKLIPELKNKLSKVONFFTLSTNV-----KQANKD 235
Db 1248 -----NTNITNSDSIKQIKETIKIONALSLTNKISKQKELEKYNVTTKTALEQLIKD 1302
QY 236 IDAAKLLITTEIAAI-----GEIKTETET-----TR 261
Db 1303 EDAKEVGTDDANNAITKKNADKNSTLBEITNATKALEDKSKLDQELTKKAEFPNNLTK 1362
QY 262 FYVDYDDLMLSLKKEAAKKNMTCNEYQKRGKKTLPFV 300
Db 1363 AKTELSDLITSSNQAPADAIS-----DAQKTLDEI 1393
RESULT 44
Q6CXQ4 PRELIMINARY; PRT; 655 AA.
AC Q6CXQ4;
DT 01-OCT-2004 (TrEMBLrel. 28, Created)
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to sp|P34237 Saccharomyces cerevisiae YKL179c.
GN ORFNames=KLLA0A06402g;
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RA GENOLEVURES;
RG Dujon B., Sherman D., Fischer G., Durrans P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marcq C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jaumiaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikoleki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382121; CAH02873.1; -- 8024192DF56618EC CRC64;
SQ SEQUENCE 655 AA; 74898 MW; 8024192DF56618EC CRC64;
Query Match 7.9%; Score 119.5; DB 2; Length 655;
Best Local Similarity 21.3%; Pred. No. 27;
Matches 69; Conservative 56; Mismatches 106; Indels 93; Gaps 13;
QY 24 LDLYNKYLQVLPWQTFDETIKELSR-----FKQEYSOQASVL--VGDIKTLMDSDQ 74
Db 91 IELYELKIEQDPDPTGILESIVEELKDGTVNDDIKAENQKLAIVSKCADIYTLKKELSD 150
QY 75 KYFEATQTV-----YEWCGVATQLLAAYILLFDEYN--EKKASQK 113
Db 151 LEQNSAKTLNRLIAKEKITSKWEKQKRNWNRRELLKQLETNNTNSILEKKIGAQV 210
QY 114 DILIKVLDDG-----ITKLNEAQKSLVSSQSFNNASGKLALDS-----QLTN 157
Db 211 D--LEQEDGDDDTVLVSGKNSAQSELLV--QEEAAQIRIMSLESNEELNSEVTKLKS 266
QY 158 DFSEKSSYFQSDVKIRKEAYAGAAAGVAGPPGLIISYSIAGVVEGKLIPELKNKLS 217
Db 267 D-BQDSIFQEKETKLNQLESENA-----KLVTCLEERKS 301
QY 218 VQNFFTLSTNVKQA-----NKDIDAACKLKT-----ETAAIGEIK--TETET 259
Db 302 LKETSQTLSQLSSALETQTYKSELETURRKLNTNYSYDEKIKQELNMMKKIEFGTSDS 361
QY 260 TRFYVDYDDLMLSLKKEAAKKNM 283
Db 362 DNENDGDDSDITSSLKHANQKLQN 385
RESULT 45
Q86TB8 PRELIMINARY; PRT; 973 AA.
AC Q86TB8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein DKFZp451D234.
GN Name=DKFZp451D234;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
```

```

RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Oeanger A.,
RA Fobo G., Han M., Wiemann S.;
RA Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL832010; CAB89917.1; -.
DR GO; GO:0005737; C:cytosolasm; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008565; P:protein transporter activity; IEA.
DR GO; GO:0006886; P:intracellular protein transport; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR002025; Armadillo.
DR InterPro; IPR006955; Uso1_p115_C.
DR InterPro; IPR006953; Uso1_p115_head.
DR Pfam; PF04871; Uso1_p115_C; 1.
DR Pfam; PF04869; Uso1_p115_head; 1.
DR PROSITE; PS0176; ARM_REPEAT; 1.
KW Hypothetical protein.
SQ SEQUENCE 973 AA; 109186 MW; 9DD0279FC2CC6BFD CRC64;

Query Match 7.9%; Score 119.5; DB 2; Length 973;
Best Local Similarity 23.7%; Pred. No. 42;
Matches 74; Conservative 51; Mismatches 136; Indels 51; Gaps 13;

QY 7 DKTEVVVKNATETADGALDLY-NKYLDQVTPWQTFDETIKELSRFKQEYSQAASVLVGD 65
Db 646 DKKEEVKVLQEHDDIVTHYKNMIREQDQLQELRQOVSTLKCNEQLQTAVTQVQS 705

QY 66 K-----TLIM-----DSQKYEATQTVYEWCGVATQLLAAYILLFDEYNEKKA 110
Db 706 QQHKDQYNLLKIQLGKDNQHGYSQAQ-----MNGIQPEEIGR---LREEIEELKRN 758

QY 111 -----AKDILIKVLDGDTKNEAKSKLLVS-----SQSFNNSGKLLALDSQ 161
Db 759 LLQSQTKEKDSMIENKSSQTSQSGTNEQSAIVSARDSEQVAELKQELATLKSQ 817

QY 162 KSSYFQSQVDKIRKEAVAGAAAGVAGFPGLIISYSIAAGVEGL-----IPEL 215
Db 818 EITKLQTEKQELLOKTEFAKSVQVQETETIATKTTD--VEGRSLALQETKELKNEI 875

QY 216 KSVQNFPTTISNTVKQANKDIDAAKJLTTETIAAIGBIKTETETTRFPVYDD 275
Db 876 KALSEERTAIKEQLDSSNSTIAI-----LQTE-----KDKLEITDTSKQEQD 925

QY 276 EAAKMMINTCNE 287
Db 926 DQDQKILSLKKNK 937

RESULT 46
Q6BUQ9 PRELIMINARY; PRT; 2042 AA.
AC Q6BUQ9
DT 01-OCT-2004 (TrEMBLrel. 28, Created)
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to sp|P25386 Saccharomyces cerevisiae YDL058w USO1.
GN ORFNames=DEHA0C09658g;
OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=4959;
RN [1]
SQ SEQUENCE FROM N.A.
RC STRAIN=CHS767;
RG GENOLVRURES;
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Dufon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boismare A., Boyer J., Catholico L., Confanier F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,

```

```

RA Pollenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Wenhene D., Tekata F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CHS767;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382135; CAG86127.1; -.
SQ SEQUENCE 2042 AA; 233328 MW; 959292DEB8EE1158 CRC64;

Query Match 7.9%; Score 119.5; DB 2; Length 2042;
Best Local Similarity 19.9%; Pred. No. 1e+02;
Matches 59; Conservative 60; Mismatches 124; Indels 53; Gaps 10;

QY 8 KTVVVKNAIETADGALDLY-NKYLDQVTPWQTFDETIKELSRFKQEYSQAASVLVGD 67
Db 1695 KSIQSLSALKSSDKAASEMTKQLENEL--QTLKDDIEEKSRKSELEEKSTLSS 1752

QY 68 L--LMDSDQKYEATQTVYEWCGVATQLLAAYILLFDEYNEKKAQAQDILIKVLD 125
Db 1753 LENKLDAMKELSEKSVIEKLSAEKHSKLSADLKEYKEFEQLEKE-----HE 1806

QY 126 KLNEAQSKLLVSSQSFNNSGKLLALDSQLTNDPSEKSSYFQSQVDKIRKEAVAGA 185
Db 1807 KKFDAEGN--IHGKMKELKSLDLSLQDDLTAAMDLLK-----SKIESLNQEL 1851

QY 186 VAGFPGLIISYSIAAGVEGLKIPELKNKLKSVNFTTISNTVKQANKDIDAAKJL 245
Db 1852 -----LSTKTTKDDKIKLTKDLESTQ-----ALKNNKELKLDLNSKEN 1895

QY 246 ---EIAAIGBIKTETETTRFPVYDDMLSLKKE-AAKMMINTCNEYQKRHGK 297
Db 1896 LKEDLNALKENETRLK-----QDLSEKANEDSLNKELESLSQSKQKNEK 1943

RESULT 47
Q9LIW7 PRELIMINARY; PRT; 895 AA.
AC Q9LIW7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Similar to an Arabidopsis thaliana chromosome BAC genomic
DE sequence.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Haing Y.C., Chow T., Chen C., Wu H., Chu M., Chao Y., Liu S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001111; BAA90502.1; -.
DR Gramene; Q9LIW7; -.
SQ SEQUENCE 895 AA; 100520 MW; BAA4F7BFD70FFAA3 CRC64;

Query Match 7.9%; Score 119; DB 2; Length 895;
Best Local Similarity 21.1%; Pred. No. 41;
Matches 74; Conservative 74; Mismatches 123; Indels 80; Gaps 18;

QY 1 MTEIVADKTVEVKNATETADGALD-----LYNKYLDQVTPWQTFDETIK----- 47
Db 268 IAEVNAEK-VEILSSSEVVRKGLLDSTAESSEKNEETEELV--KNLESEVSVLKGL 324

QY 48 SRFKQEYSQAASVLVGDITKLLMDSDQKYEATQTVVW---CGVATQLLAAYILL 104
Db 325 ARTIEERLAETKLIBELKSEVADAKKAESARQLFEWKHKAGLLEMELEA-VTLSD 383

```

```

QY 105 -NEKASAKQDILIKVLDDGKITKLENAQKSLVSSQSFNNASGKLLALD---SQTNDPFS 160
Db 384 KGESLASTTBE-----LGKIQSALQDRESIEYVLKGTTALETAEVARLLADV 431
QY 161 EKSYF-----QSOVDKIRKEAYAGAAAGVAGPFLIISYSIAAGV-VEG--- 205
Db 432 ESNEQFDASQOEVLGQTTIDVLNKLKLEAAEAASEA-----LNNEKAANVKIEGLTEE 485
QY 206 --KLIPEL-----KNLKSQVNFPTLNTVKQANKOIDAAKLKLTTE-----IAAI 250
Db 486 NVKLISELNTDREKEKEKRAVEDLTAALS---EESDKAEAEHVELSKEDDHEHALAQI 542
QY 251 GEIKTETETTR-----FVYVDDMLSLKLEAAKMINTCNEYQKR-HGKKT 296
Db 543 GDLKMAKSTKSEYEVMLDEANDYITCLRNKVDKLEAVNKNYEECESKET 593

RESULT 48
Q7RELO PRELIMINARY; PRT; 2664 AA.
AC Q7RELO
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Rhoetry protein (Fragment).
GN Name=PY05054;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Koolij T.W., Perleeta M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shalom S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoalibi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001582; EAA17018.1; -.
DR InterPro; IPR006499; ReticulocyteBP.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
FT NON TER 2664
SQ SEQUENCE 2664 AA; 312670 MW; FDE72DB05743F4AB CRC64;

Query Match 7.9%; Score 119; DB 2; Length 2664;
Best Local Similarity 22.1%; Pred. No. 1.5e+02;
Matches 90; Conservative 50; Mismatches 122; Indels 146; Gaps 21;

QY 1 MTEIVADKTVVKNQNAETADGALDLYNKYLDQVLPWQTFDETKE----- 46
Db 819 MTKI-STKDEILKNINEVK-----HMKDNFLDKNIYTNFDIYKENVDSHDQFTELAK 873
QY 47 -----LSRFKQEYSQAASVLVGDITKLLMDSQDKYEATQV-----VEMCGVAT 91
Db 874 KIKTDVASKFNEYEQ-----MLGSKALITQTKSIEEYQNTLKKANEYQLCDNTT 929
QY 92 QLLAAAYILLDFEYNE-----KKASAQKDI-----LIKVLDDGI---TKLNE 129
Db 930 ELIKFNFKNELSEILKNKINIESINNTSIEKSYAENFINTLNKTKELDDIFKDAASLNE 989
QY 130 AQK-----SLLVSSQSFNNASGKLLALDLSQTLNDFSEKSSYFOSQV 170

```

```

Db 990 HKTKTEELKTYFNDLKENLGKSNSTLSQOFN---CK-----EKLFDNIQKVNIDINKNI 1041
QY 171 DKIRKEAYAGAAAGVAGPFLIISYSIAAGV---VEGKL-----IPE 210
Db 1042 SNIEKEIYSSI-----HNINEDIENETEKNIKSLNTQVFEKVKTNVANLNE 1087
QY 211 LKNKLSQVNF-----FTTSLNTVKQ---ANKDIDAAKKLKLTTEI-----AAIG 251
Db 1088 IREKLAQ-QYNODFEKEKNKYTDENWKIKDDVTTLNKKIDKNITKL-TIEKNSESYS 1145
QY 252 EIKTETETTRFYVDYDMLSLKLEAAKMINTCNEYQKRHGKKTLPFE 299
Db 1146 EIKTQDKSEKTYD-KTIYKEEPEKEKKIKNIITKIDR---KKNIFE 1189

RESULT 49
Q7P7E4 PRELIMINARY; PRT; 335 AA.
AC Q7P7E4
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12).
GN Name=FNVL630;
OS Fusobacterium nucleatum subsp. vincentii ATCC 49256.
OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=209882;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49256;
RA Karpatrial V., Ivanova N., Anderson I., Resnik G., Bhattacharyya A.,
RA Gardner W.L., Mikhailova N., Larsen N., D'Souza M., Walunas T.,
RA Haselkorn R., Overbeek R., Kyrpides N.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the glyceraldehyde-3-phosphate
CC dehydrogenase family.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABF01000017; EAA24760.1; -.
DR GO; GO:0004365; F:glyceraldehyde-3-phosphate dehydrogenase (p. . .; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006006; P:glucose metabolism; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR006424; GAPDH-I.
DR InterPro; IPR011596; GAPDH like.
DR InterPro; IPR000173; GAP_dhrogenase.
DR Pfam; PF02800; Gp_dh_C; 1.
DR Pfam; PF00044; Gp_dh_N; 1.
DR PRINTS; PR00078; G3PDHGRNASE.
DR ProDom; PD007761; GAPDH like; 1.
DR TIGRFAMs; TIGR01534; GAPDH-I; 1.
DR PROSITE; PS00071; GAPDH; 1.
KW Oxidoreductase.
SQ SEQUENCE 335 AA; 35833 MW; 5089AB90E7D16B61 CRC64;

Query Match 7.9%; Score 118.5; DB 2; Length 335;
Best Local Similarity 22.3%; Pred. No. 14;
Matches 73; Conservative 42; Mismatches 110; Indels 103; Gaps 16;

QY 7 DKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFD-ETIKELSRF--KQYSQA----- 57
Db 55 DGTIEVTDNGFVVGSDISKVFAPKANPEELPWGLDGLDVLECTGFTGFTSKEKAHAKAGA 114
QY 58 -----ASVLVGDITKLLMDSQDKYEATQVTCVWCGVATQLLAAYILLDFEYNEKASQA 112
Db 115 KKVVISAPATGDKLTIVYVNVNVDVGTETVISGASCTTNCCLAP----- 158
QY 113 KDILIKVLDD--GITKLENAQKSLVSSQSFNNASGKLLALDLSQTLNDFSEKSSYFOSQV 170
Db 159 ---MAVLNDKFGII-----EGLMTTIHAYTN-----DQNTLDAPHKKG----- 194

```

|    |     |   |     |
|----|-----|---|-----|
| Qy | 171 | DKIRKEAVAGAAAGVVGFGGLIISYSTAAGVVGKLIPELKNKLK-----SVQ          | 219 |
| Db | 195 | DLRRARA---AAENIVPNTTG-----AAKAICGLVIPELKGLDGAORVPIGTSIT       | 243 |
| Qy | 220 | NFFFTLSN--TVQKANKIDIAKLK-----LTTTIAAI--GEIKTETETTRFVD-        | 265 |
| Db | 244 | ELVTVLGKDVTVDEYNAANKAAASNEFQYETPEPLVSSDIIGISFGSLFDATQIKVLTVDG | 303 |
| Qy | 266 | -----YDDLK--LSLLKEAAKKMI                                      | 282 |
| Db | 304 | KOLIKVTVAWYDNEMSYTQSILRTLUKKFV                                | 331 |

## RESULT 50

|        |  |              |      |         |
|--------|--|--------------|------|---------|
| Q895W1 | Q895W1   | PRELIMINARY; | PRT; | 495 AA. |
| ID     | Q895W1   |              |      |         |
| AC     | Q895W1;  |              |      |         |
| DT     | 01-JUN-2003 (TrEMBLrel. 24, Created)                               |              |      |         |
| DT     | 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)                  |              |      |         |
| DT     | 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)                |              |      |         |
| DE     | Methyl-accepting chemotaxis protein.                               |              |      |         |
| GN     | OrderedlocusNames=CTC01155;  |              |      |         |
| OS     | Clostridium tectani.   |              |      |         |
| OC     | Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;   |              |      |         |
| OC     | Clostridium.   |              |      |         |
| OX     | NCBI_TaxID=1513;   |              |      |         |
| RN     | [1]  |              |      |         |
| RP     | SEQUENCE FROM N.A.   |              |      |         |
| RC     | STRAIN=Massachusetts / E98;  |              |      |         |
| RX     | MEDLINE=22457235; PubMed=12552129;                                 |              |      |         |
| RA     | Brueggemann H., Baumer S., Fricke W.F., Wierse A., Liesegang H.,   |              |      |         |
| RA     | Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,     |              |      |         |
| RA     | Gottschalk G.;   |              |      |         |
| RT     | "The genome sequence of Clostridium tetani, the causative agent of |              |      |         |
| RT     | tetanus disease."  |              |      |         |
| RL     | Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321 (2003).                |              |      |         |
| DR     | ENBL; AEO15940; AAQ35729.1; --                                     |              |      |         |
| DR     | GO; GO:001620; C:membrane; IEA.                                    |              |      |         |
| DR     | GO; GO:0004871; F:signal transducer activity; IEA.                 |              |      |         |
| DR     | GO; GO:0006935; P:chemotaxis; IEA.                                 |              |      |         |
| DR     | GO; GO:0007165; P:signal transduction; IEA.                        |              |      |         |
| DR     | InterPro; IPR004089; Chmtaxis transd.                              |              |      |         |
| DR     | InterPro; IPR004090; Mc chemotaxis.                                |              |      |         |
| DR     | Pfam; PF000015; MCPsignal; 1                                       |              |      |         |
| DR     | PRINTS; PR00260; CHEMTRNSDUCR.                                     |              |      |         |
| DR     | SMART; SMO0283; MA; 1.   |              |      |         |
| DR     | PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.                        |              |      |         |
| DR     | Complete proteome.   |              |      |         |
| SW     | SEQUENCE 495 AA.   |              |      |         |
| SQ     | SEQUENCE 495 AA: 55854 MW: 4167C6BD598557CB CRC64;                 |              |      |         |

|                       |        |               |       |                                     |
|-----------------------|--------|---------------|-------|-------------------------------------|
| Query Match           | 7.8%;  | Score 118;    | DB 2; | Length 495;                         |
| Best Local Similarity | 20.6%; | Pred. No. 24; |       |                                     |
| Matches               | 63;    | Conservative  | 59;   | Mismatches 122; Indels 62; Gaps 12; |

  

|    |     |  |     |
|----|-----|--|-----|
| Qy | 12  | VVKNATETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQEVYQAAASVLVGDIKTL-- | 69  |
| Db | 18  | IIKYVERLEG-----NKTL-----PKVEYPIHVKLKNYKGLFSN-EGIMSSSAKTLDDI  | 68  |
| Qy | 70  | -MDSQDKYFATCTVYECGWATOLLAAYILLFDEYNEKKAQAQKOLIKV--LDDGITK    | 126 |
| Db | 69  | NASLSDFDVQWSSISYELIDFAKEM-----SELSESNLAVVEITATSMNQVNHITIED   | 120 |
| Qy | 127 | LNIRAQSKLLVSSQSFNNASGKLLA-----LSQLTNDFSEKSSYFQSQVDKIRKAYAY   | 179 |
| Db | 121 | TSKLTEDLSISSEKLIBEENHKSIAEIEDINHLKEEVNANNIMISSIEELVEVANKVSDI | 180 |
| Qy | 180 | GAAGVAGVAGFGLII-----SYSTIAGVVEGKLIPELKNKLKSVQNFPTT           | 224 |
| Db | 181 | VEGVGATADQTNLLIALNASTEARAGBGKGFAVVAQEIR-KLADDTKGSLOQNRNFPVN  | 239 |
| Qy | 225 | LSNTVKQANKDID---AAKLKLTATTEIAAIGSEIKTETTFYVDYDIDMLLSLKEAAKQM | 281 |

```

Db      240  IONTREGKKSMNDTISSTEKMSKKIDAI-----TYTTKSNVD-----MLEDVSRS 285
Qy      282  INTCNE 287
        |  |  |
Db      286  IYTIME 291

```

Search completed: January 5, 2005, 10:58:28  
Job time : 59.3174 secs